

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:07:18 ; Search time 1381.67 Seconds
(without alignments)
-3065.678 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgttcaaaaataccaata.....tcagggaagtcctcgtaaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_bal.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	365.6	26.2	1961	2	PAU32853	U32853 Pseudomonas
2	365.6	26.2	4587	2	PAU49151	U49151 Pseudomonas
3	274.8	19.7	2027	1	AVU30799	U30799 Azotobacter
4	212.4	15.2	1974	2	RMU31512	U31512 Rhizobium m
5	211.8	15.2	2009	1	BUU07352	U07352 Brucella ab
6	197	14.1	1905	1	BRUHTRAH	L09274 Brucella ab
7	196	14.1	2714	2	ECU32495	U32495 Escherichia
8	196	14.1	3156	2	ECU15661	U15661 Escherichia
9	196	14.1	10713	2	AE000402	AE000402 Escherich
10	196	14.1	110000	2	ECOUW67_1	Continuation (2 of
11	186.6	13.4	1191	2	PAU29172	U29172 Pseudomonas
12	178	12.8	10199	2	AE000125	AE000125 Escherich
13	178	12.8	82727	1	EC082K	D26562 Escherichia
14	178	12.8	123171	1	ECU70214	U70214 Escherichia
15	176.4	12.6	1855	1	ECHTRA	X12457 Escherichia
16	176.4	12.6	1855	1	ECOHTRAA	M36536 E.coli htra
17	175.6	12.6	1980	1	STSPHSPG	X54548 S.typhimuri
18	175.6	12.6	1980	5	AR037634	AR037634 Sequence
19	175.2	12.6	1980	5	A18802	A18802 htra gene.
20	164.6	11.8	4151	1	RCFDXE	Y11304 R.capsulatu
21	164.4	11.8	1800	1	D78376	D78376 Versinia en
22	164	11.8	110000	44	AC009220_1	Continuation (2 of
23	157.2	11.3	1613	1	BAU07351	U07351 Brucella ab
24	154.2	11.1	1791	5	AR000272	AR000272 Sequence
25	154.2	11.1	1791	5	I51141	I51141 Sequence 7
26	154.2	11.1	1791	5	I78681	I78681 Sequence 7
27	154.2	11.1	1889	1	ROCHTRA	L20127 Rochalimaea
28	149.6	10.7	11201	2	AE001732	AE001732 Thermotog
29	148.2	10.6	11881	2	AE001355	AE001355 Chlamydia
30	147.6	10.6	2778	1	BJY13616	Y13616 Bradyrhizob
31	139.6	10.0	33773	41	AC008862	AC008862 Homo sapi
32	139.4	10.0	1509	2	AF018152	AF018152 Haemophil
33	139.4	10.0	2894	5	AR034125	AR034125 Sequence
34	139.4	10.0	2894	5	I19541	I19541 Sequence 1
35	139.4	10.0	2894	5	I60405	I60405 Sequence 1
36	139.4	10.0	2894	5	I64380	I64380 Sequence 1
37	138.8	9.9	1506	1	YEHTRA	X94153 Y. enterocol
38	138.4	9.9	10477	2	AE001474	AE001474 Haemophil
39	137.8	9.9	1314	2	AF018151	AF018151 Haemophil
40	137.8	9.9	12480	2	U32805	U32805 Haemophilus
41	136.2	9.8	1611	5	AR061048	AR061048 Sequence
42	136.2	9.8	1611	5	I70313	I70313 Sequence 1
43	136.2	9.8	1611	5	I89266	I89266 Sequence 1
44	133.4	9.6	13485	2	AE001678	AE001678 Chlamydia
45	131.6	9.4	42688	41	AC009092	AC009092 Homo sapi

ALIGNMENTS

RESULT 1
PAU32853 PAU32853 1961 bp DNA BCT 06-MAR-1996
LOCUS Pseudomonas aeruginosa mucC and mucD genes, complete cds.
DEFINITION U32853
ACCESSION U32853
VERSION U32853.1 GI:1184682
KEYWORDS

TITLE	Mechanism of conversion to mucoidy in <i>Pseudomonas aeruginosa</i> infecting cystic fibrosis patients	-10_signal	sigmaE promoters of <i>E. coli</i> in both the -10 and -35 sequences"
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8377-8381 (1993)	mRNA	411. .415
MEDLINE	93391358		/note="P1 promoter"
REFERENCE	4 (bases 252 to 488)		421
AUTHORS	Martin,D.W., Schurr,M.J., Yu,H. and Deretic,V.		/note="transcription start site for promoter P1; 3' end is
TITLE	Analysis of promoters controlled by the putative sigma factor AlgU regulating conversion to mucoidy in <i>Pseudomonas aeruginosa</i> : relationship to sigma E and stress response	gene	unknown"
JOURNAL	J. Bacteriol. 176 (21), 6688-6696 (1994)		474. .1055
MEDLINE	95050230	gene	/gene="algU"
REFERENCE	5 (sites)	CDS	474. .1055
AUTHORS	Yu,H., Schurr,M.J. and Deretic,V.		/gene="algU"
TITLE	Functional equivalence of <i>Escherichia coli</i> sigma E and <i>Pseudomonas aeruginosa</i> AlgU: <i>E. coli</i> rpoE restores mucoidy and reduces sensitivity to reactive oxygen intermediates in algU mutants of <i>P. aeruginosa</i>		/codon_start=1
JOURNAL	J. Bacteriol. 177 (11), 3259-3268 (1995)		/transl_table=11
MEDLINE	95286510		/function="conversion to mucoidy; required for algD
REFERENCE	6 (bases 1 to 252)		transcription"
AUTHORS	Schurr,M.J., Yu,H., Boucher,J.C., Hibler,N.S. and Deretic,V.		/product="sigma factor"
TITLE	Multiple promoters and induction by heat shock of the gene encoding the alternative sigma factor AlgU (sigma E) which controls mucoidy in cystic fibrosis isolates of <i>Pseudomonas aeruginosa</i>	gene	/protein_id="AAC43714.1"
JOURNAL	J. Bacteriol. 177 (19), 5670-5679 (1995)		/db_xref="GI:1173502"
MEDLINE	96032406	CDS	/translation="MLTQEQDQQLVERVQGRKRAFDLLVLKYQHKILGLIVRFVHDA
REFERENCE	7 (bases 1 to 4587)		QEQDVAQAFIKAYRAGNFRGDSAFYTWLYRTIAINKHNLVARGRPDSDVTAE
AUTHORS	Boucher,J.C., Martinez-Salazar,J., Schurr,M.J., Mudd,M.H., Yu,H. and Deretic,V.		DAEFEGDHAKDIESPERAMLRDEIEATVHTQIQQLPEDLRTALTTLREFEGLSYDEI
TITLE	Two distinct loci affecting conversion to mucoidy in <i>Pseudomonas aeruginosa</i> in cystic fibrosis encode homologs of the serine protease HtrA		ATVNCQCPVGTVRSRIFRAREIDKALQPLLEA"
JOURNAL	J. Bacteriol. 178 (2), 511-523 (1996)		1087. .1671
MEDLINE	96134987		/gene="mucA"
REFERENCE	8 (bases 1 to 4587)		1087. .1671
AUTHORS	Boucher,J.C., Martin,D.W., Schurr,M.J., Deretic,V., Yu,H., Mudd,M. and Martinez-Salazar,J.		/gene="mucA"
TITLE	Direct Submission		/codon_start=1
JOURNAL	Submitted (09-FEB-1996) John C. Boucher, Microbiology, U.T. HSC San Antonio, 7703 Floyd Curl Dr., San Antonio, TX 78284, USA	gene	/transl_table=11
COMMENT	On Mar 11, 1996 this sequence version replaced gi:1173501.		/product="MucA"
FEATURES	Location/Qualifiers	CDS	/protein_id="AAC43715.1"
source	1. .4587		/db_xref="GI:1173503"
	/organism="Pseudomonas aeruginosa"		/translation="MSREALQETLSAVMNEADELELRVLAAACGEDAEELRSTWRSYQ
	/strain="PAO1"		LARSVMHREPTLPKLDIAAASAAALADEAAPPKAEKGPWVMVGRILAAASVTLAVLAG
	/db_xref="taxon:287"		VRLYNNDALPQMAOQGTTPQIALPOVKGPVAVLAGYSEEQAGPQVITNSSSDTWRHE
	/map="67.5"		QRPLVLRQHVQQQSAVSGSTESALPYARAASLENR"
mRNA	22		1680. .2630
	/note="transcription start site for promoter P5; 3' end is unknown"		/gene="mucB"
	124		1680. .2630
mRNA	124		/gene="mucB"
	/note="transcription start site for promoter P4; 3' end is unknown"		/codon_start=1
promoter	204. .238		/transl_table=11
-35_signal	204. .209		/product="MucC"
-10_signal	226. .230		/codon_start=1
mRNA	235		/transl_table=11
	/note="AlgU-dependent transcription start site for promoter P3; 3' end is unknown"		/product="MucC"
mRNA	238		/protein_id="AAC43717.1"
	/note="AlgU-dependent transcription start site for promoter P3; 3' end is unknown"		/db_xref="GI:1220193"
mRNA	310		/translation="MIEQGRVVAEPGAVVETVRRSTCCSCSANACGGQGLMORLGL
	unknown"		VGAGRVRVALSDLSLRVGDVNLGHEDLLRLASVLFPLPLGFFFAVLAALTRAGL
-35_signal	389. .394		VEPLIIVSGIAGLLAALLVRRHARRHADDPASQPVVLRALISGPSDSA"
promoter	389. .421		3122. .4546
	/note="P1 promoter, the first of two AlgU-dependent promoters; AlgU-dependent promoters strongly resemble the		/gene="mucD"
			3122. .4546
			/gene="mucD"
			/codon_start=1
			/transl_table=11
			/product="MucD"
			/protein_id="AAC43718.1"
			/db_xref="GI:1220194"
			/translation="MHTLKSCMAAMVALLALSMTARAELPDTFTPLVEQASPAVYNI
			STRQKLPDRAMARGQLSIDPLEGLPMPFRDFLEKSIPOVPRNPRCQQRQAEQSLSGSFI

ISNDGYILTNHVVADAILVRLSDRSEHKAKLIGADPRSDVAVLKIENKLPKLI GDSNKLKGEWLAIGSPGFDHSVTAGIVSAKGRSLPNESYVFFQTQDVAIPNGSG GELLNQGVEVINGSIQIFTRSGFGMLGSAIPIDVALNVADQLKAKGKVRGWLGVVI QEVNKDLAESFGLDKPSGALVAQLVEDGPAAGKGLQVGDVILSLNGQKINRSLADLPHL VGNMKPGDKINLDVIRNGRKSLSMVAGSLPDDDEETIASMAGPACRSSNRGLGVTVAD LTAQRKSLIDIOGGVVIKEVQDGPAAVIGLRPGDVITHLNDKAVTSTKVPADVAKALP KNRSVMRLVROGRASFITFKLAE"										
BASE COUNT	811 a	1447 c	1515 g	814 t						
ORIGIN										
Query Match	26.2%	Score 365.6	DB 2	Length 4587						
Best Local Similarity	68.0%	Pred. No. 5e-65								
Matches 525	Conservative 0	Mismatches 244	Indels 3	Gaps 1						
QY	374	togggttcgggttcctcatcagcaaaacggtcctacatctcgtgaccataaccacgtcgttg	433							
Db	3411	TGGGTTCCGGCTTCATCTCCAACGACGGCTACATCCTCACCAACAATCACGTCGTGG	3470							
QY	434	cgggtatgggcagtatcaaaagtctgctcaacgacagcgcggaatataccgccaactca	493							
Db	3471	CCGATCCGACGAGATCTCTGCTGCGCTGTCCGACCGTAGCGAGCACAAAGGCCAAGTTGA	3530							
QY	494	tcgggttcgagtgccaatccgatgctgcctctctgaaatcgcagcgaacgaagctac	553							
Db	3531	TCGGCGGACCCGCGACGACGCTGGCGGTGCTGAAGATCGAGG---CGAAGAACCTGC	3587							
QY	554	cggtgcataaaatcggcaatcccaaaatttgaacccggcggaatgggtcgtgccatcg	613							
Db	3588	CGACCTGAAACTGGGCGATTCGAACAAGCTGAAAGTGGCGAATGGGTCTCGGCCATCG	3647							
QY	614	ggcgcccttcggttttgacaacagcgtgaccgcgcatcgtgctccgcgaagggcgaa	673							
Db	3648	GTTCCGCGTTCGGCTTCGATCACTCGGTCAACCGCGGTATCGTCAGTGCCCAAGGTCGTA	3707							
QY	674	gctgcgccaaacgaagatcacaccccttcatacaaacgacgttgccataatccggcca	733							
Db	3708	GCCTGCGCAACGAGAGCTAGCTACCCTTCATCCAGACCGACGTCGCGATCAACCCGGGTA	3767							
QY	734	attccggcgccgctgttcaacttaaaagagacaggtcgtcggcatcaattcgcgaatat	793							
Db	3768	ACTCCGCGGTCCGCTGCTGAACCTGCGAGGCGAAGTGTGCGGCATCAACTCGCAGATCT	3827							
QY	794	acagcgcgagcgcggtatcatatgggcatctccttggcatcccgatcgagcttgcctatga	853							
Db	3828	TCACCGTTCGCGGGCTTCATATGGCCCTGCTCCTTCGCCATTCGATCGATGTCGCGTGA	3887							
QY	854	atgtcgcgcaaacagctgaaacacaccgcgaagtcaccaacgcggcaactggcgtgatta	913							
Db	3888	ACGTCCCGACCAAGTTGAAGAAGCCGCAAGGTCACTCGCGGTGCTGGCGGTGGTGA	3947							
QY	914	ttcagaagaatatactacggttttggcacagtcgttctcgtctcgtgataaaagcagcgccat	973							
Db	3948	TCAGGAAGTGAACAAGGATCTCGCGAGTCTTCGCGCTTCGACAAAGCCGTCGCGCGC	4007							
QY	974	tgaattcccaaaatccttcccgagccccgcagaaacgctgcgcctcagcgggcgacaca	1033							
Db	4008	TGTTGCGCAACTGGTGGAAAGACGGCCCGCGGCAAGGGTGGCTTCAGGTGGGCGATG	4067							
QY	1034	tcgtcctcagctcgcagcgcgagaaatacgtttcttcgcggaccttcctcgatcgctg	1093							
Db	4068	TGATCCTCAGCCTGAACGGCCAGTCGATCAACGAGTCCGCGACCTGCCGACCTGGTGG	4127							
QY	1094	ggccattacgcgggaagaagaagtcagctcgcgtatggcgcaaggcgga 1145								
Db	4128	GCAACATGAAGCGCGGCGACAAGATCAACCTGGACGTGATTCGCAACGGGCCA 4179								
RESULT	3									
AVU30799	2027 bp	DNA	BCT	31-MAY-1996						
LOCUS	Azotobacter vinelandii	MucC (mucC) and MucD (mucD) genes, complete cds.								

ACCESSION	U30799										
VERSION	U30799.1	GI:1345102									
KEYWORDS	Azotobacter vinelandii.										
SOURCE	Azotobacter vinelandii										
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae; Azotobacter.										
REFERENCE	1 (bases 1 to 2027)										
AUTHORS	Martinez-Salazar,J.M., Moreno,S., Najera,R., Boucher,J.C., Espin,G., Soberon-Chavez,G. and Deretic,V.										
TITLE	Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA, MucB, MucC, and MucD in Azotobacter vinelandii and evaluation of their roles in alginate biosynthesis										
JOURNAL	J. Bacteriol.	178 (7),	1800-1808	(1996)							
MEDLINE	96178940										
REFERENCE	2 (bases 1 to 2027)										
AUTHORS	Martinez-Salazar,J.M., Moreno,S., Najera,R., Boucher,C., Espin,G., Soberon-Chavez,G. and Deretic,V.										
TITLE	Direct Submission										
JOURNAL	Submitted (30-JUN-1995) Jaime M. Martinez-Salazar, Microbiology, UTHSCSA, 7703 Floyd Curl Drive, San Antonio, TX 78284, USA										
FEATURES	Location/Qualifiers										
source	1..2027										
	/organism="Azotobacter vinelandii"										
	/strain="UW 136"										
	/db_xref="taxon:354"										
gene	1..471										
	/gene="mucC"										
CDS	1..471										
	/gene="mucC"										
	/note="similar to Photobacterium sp. ORF4 product, encoded by GenBank Accession Number L41667"										
	/codon_start=1										
	/transl_table=11										
	/product="MucC"										
	/protein_id="AAB01512.1"										
	/db_xref="GI:1345103"										
	/translation="MTEEPGRVVAIERAVNIEALDEAACPGCTASVCGGRMSFDWLETPRRRRIRALSVNPLRVGDAVVVGIREDFLRLSTLAIYLLPLGLFTCTIAIOQLGFGSLMIPAGFLGVSVWAVRRISRRMNDPTLQPVVLRALPGASEGVQRAAF"										
CDS	483..1904										
	/note="similar to Brucella abortus HtrA protein encoded by GenBank Accession Number U07352"										
	/codon_start=1										
	/transl_table=11										
	/function="serine protease"										
	/product="MucD"										
	/protein_id="AAB01513.1"										
	/db_xref="GI:1345104"										
	/translation="MSKVGFKSFASVLAGALLQGSFLVQAQLPEFTSLVEASPAVV NISTROKLPDRSTVQGLPDLGLPLREFLESLIPOLPTPDNGRQREAHSLGSGFI ISPDGYLTNNHVVADAEIIVRLSDRSELAELVGADPLTRCSFEVQGGFSPHSQGTGYPIESRGMGSGHRPFRPSRDCGHCHGAKPAERLVVPIOTDVAINPNSGIVGFLPDLGRVIGINSQIFTRSGGFMGLSFAIPLEVAMGVADQLKATGVARGMLGVII QEVNKDLAESFGLDRPAGALVAQVLEDGPADKGGQLQGVLDVILSDGHPIVMSADLPHLVGGILKPGAAANLEVVDRKRNRIATVIGALPEEGNVQPSIAGTQEQRNRGLGVTVTEL TAEQKSLDLKGGVYIRELVNGPAALIGLRPGDVVYTHLNNQPIDSAKTFAEVAGALPKGRSVSMRVLQRGRASFITFKLAE"										
BASE COUNT	366 a	529 c	658 g	474 t							
ORIGIN											
Query Match	19.7%; Score 274.8; DB 1; Length 2027;										
Best Local Similarity	58.9%; Pred. No. 1.5e-46;										
Matches 527; Conservative	0; Mismatches 362; Indels 5; Gaps 3;										
QY	295	ttctacgaatttttcaaacgcctcgtcccgacaacatcccgaaatcccccaagaagca	354								
Db	693	TTCAGGAGTTCCTGGACGCGACGATTCGCGCAACTCCGCGTACTCCGGATAACGGCGG	752								
QY	355	gatgacgcgcatgaaattcgttcgggttcctcatcagcaaaaacggctacatcctg	414								
Db	753	CAGCGTAGGCGCACTCCTCGGCTCGGGTTTCATCATTTCTCCAGATGGCTATGTCTA	812								


```

|||||
Db 1183 CCAGCCGGTACGAGGATATCGCGAGTCTGCTCGCCCTTTCCGAGCGAAGCGGCCCT 1242
QY 975 gattgccaaatcttcccccagccccgcagaaagtcgcccgcctgcaggcgagacat 1034
Db 1243 CGTCTGAGCTCAGCGGGCTCGCCGGTGAAAAGCGCGGATCAAGAAGCGGACGT 1302
QY 1035 cgtctcagctcagcgcgagagaaatcgtttctccggcgacattcccgctcatgctcgg 1094
Db 1303 GGTGAGCGCATCTAACGGCGAGCCGGTCAAGAGATCCGGGGACCTCGCACGTGCGGTTGC 1362
QY 1095 cgcattacgcgggaaaaagtcagcctcggcgatggtgcgaaagcgcaagaaatcac 1154
Db 1363 GGCCCTCGCCCGCGCTCCATGTCAGAGGTTACCCCTGTGGCGCAGCGCAATTCGGAAC 1422
QY 1155 aatcaaaagccagtgggca---acgcgcgcagcagcaccgcgcgatcatccaaacaga 1211
Db 1423 GGTCAACCTCAGATCGGCAGCTGCCGAGCGATGCCAAGGAACCGCACCGCGACCGG 1482
QY 1212 tgaagcccttacaccgaacgaatcgttacgtttctcgttcggaatccgcaaggattac 1271
Db 1483 TGAAGCGCAGCCCGACGAAGGCGCAGGCGAGGAGGCGCTGGCGCCTCGCTCGATCCGGATCGGA 1542
QY 1272 cct---tcagacacataccgacagcagcggaacacacctcgtcgttacggtttccga 1328
Db 1543 GGTACCCCGTCGAGGAGCGCAAGGGCGTCACGATCGCTCGCTCGATCCGGATCGGA 1602
QY 1329 cgcggcgaacgcgcagcgttaagcagcagcggaacgaataccttagcgtca 1378
Db 1603 CGCCGGCGATCGC---GGCCTGAAGGAGCGGAGAGATCGTCTCCGTCA 1649

RESULT 5
BAU07352 2009 bp DNA BCT 22-JUL-1995
LOCUS Brucella abortus htrA gene, complete cds.
DEFINITION U07352
ACCESSION U07352
VERSION U07352.1 GI:497156
KEYWORDS
SOURCE Brucella abortus.
ORGANISM Brucella abortus
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
REFERENCE 1 (bases 1 to 2009)
AUTHORS Tatum,F.M., Chevillie,N.F. and Morfitt,D.
TITLE Cloning, characterization and construction of htrA and htrA-like
mutants of Brucella abortus and their survival in BALB/c mice
JOURNAL Microb. Pathog. 17 (1), 23-36 (1994)
MEDLINE 9516590
REFERENCE 2 (bases 1 to 2009)
AUTHORS Tatum,F.M.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-1994) Fred M. Tatum, Brucellosis Unit, National
Animal Disease Center/ARS/USDA, P.O.Box 70, 2300 Dayton Road, Ames,
IA 50010-0070, USA
FEATURES
source Location/Qualifiers
1..2009
/organism="Brucella abortus"
/strain="2308"
/db_xref="taxon:235"
275..1816
/gene="htrA"
275..1816
/gene="htrA"
/codon_start=1
/transl_table=11
/protein_id="AA070164.1"
/db_xref="GI:497157"
/translation="MSRARISNYRKGVAVALSALAGAFVVTGPLGALNEARAEAVH
VTTPPQAGADLVKRVPAVVSVRKKDVQETSNRGPQFFGPGFDLPDGHPLKRF
RDFCEPRGDSRSNRKANKRPHRPPVQAGSGFVISEDGYVVTNNHVVSDGDAY
TVVLDDGTDLDAKLIGADPRDLAVLNAPKRFVYVAFGDDNKRVGDMVAVGNP
FGLGGTVTSGIVSARGRDIGAGPYDDFIQIDAAYVKNKSGGPAFDLSGEVIGINTAIF"
gene
CDS

```

```

SPSGSGVGTAFAPSPSTAKQVVDOLIKGVSBERGWIGVOIPYTKDIAASGLAEKKG
AIVASPDQDGPAAKAGIKAGDVIITAVNGETVQDPRDLARKYANIAEGEKAALTVWRKN
XAEENVTIAAMPNDKSGSQSNDNDGGGETLDSYGLTVVPSDEGKGVVTVDDPD
SDAADRGIRSGDVIIVSVNNQTVKTAGDINKAITAAEKSGRKAYLLQLQSNDSREVAL
PINOR"
BASE COUNT 420 a 582 c 602 g 405 t
ORIGIN
Query Match 15.2%; Score 211.8; DB 1; Length 2009;
Best Local Similarity 54.9%; Pred. No. 1e-33;
Matches 439; Conservative 0; Mismatches 357; Indels 3; Gaps 1;
QY 376 ggttcgggttctatcatcagcaaaacggtctacatcctgaccataatccacgtgctgccc 435
Db 680 GGATCGGGCTTCGTCAATTTCCGAAGATGGCTATGCTGTTTACCAACAACCATCGTTTCC 739
QY 436 ggtatggcgagtcatacaagtcctgctcaacgcaagcgcaatataccgcaaaactcatc 495
Db 740 GATGGCGATGCCCTATACCGTCTGCTGGACGACGGACCCGAACCTTGATGCCAAGCTGAT 799
QY 496 ggttcggatgtccaatccgatgtcgcctctctgaaatcgacgcaacgga---agagcta 552
Db 800 GGTGCTGATCCGCGCACGAGACCTGGCGTGTTCGAAATCAACGCTCCGAAGCGCAAGTTC 859
QY 553 cccgtcgtcaaaatcggcaatcccaaaatttgaaccggggggaatgggtcgctgcccac 612
Db 860 GTCATATGTCGCTTTGGCGACGATATAAGGTGCGCTCGGTGATTTGGGTGTCGCAAGT 919
QY 613 ggcgcgccttcggctttgacaacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 672
Db 920 GGCATTCGTTGGCCCTTGGTGACGGTGACTTCGGGTATCGTTTCGGCCCGTGGCGGA 979
QY 673 agcctgccaaagaaagtcacacaccttcctccaaacgcgcttgccatcaatcccgggc 732
Db 980 GACATCGCGCAGCGCCCTATGACGATTTTCATCCAGATTGATCGCGCGTGAAACAAGGCG 1039
QY 733 aattccggcgcccgctgttcaactaaaaggacaggtcgccgcatcaatcgcaaatc 792
Db 1040 AATTCGGTGGTCCGGCTTTTCGACCTTTCCGGCGAAGTCATCGGCATCAACACCGCCAT 1099
QY 793 tacagcgcgagcgcggtattcatggcattctcttgccatcccgattgacgttgccatg 852
Db 1100 TTCTCCCATCTGCGCGCAGCGTGGGCATGTGCTTTTGCATATTCGCTCTCCACGCCCAAG 1159
QY 853 aatgtgcgcgaacagctgaaaaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 912
Db 1160 CAGGTGCTGGCAGCAGCTCATCAAGAAGGGTTCCGTCGAGCGCGCTGGATCGGGGTCAG 1219
QY 913 attcagaagatctctacagtttgccacagctcgttcggtctggataaagcagcgcgca 972
Db 1220 ATACAGCCTGTGACCAAGGATATCGCCGCTTCGCTCGGTCTTCGCGAGGAAAAGGGTGC 1279
QY 973 ttgattgccaaatcctcccgccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1032
Db 1280 ATCGTGGCTCGCCGAGGATGATGGCCCGCTGCCAAGGCTGGCATCAAGCGGGCGAT 1339
QY 1033 atcgtcctcagcctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1092
Db 1340 GTGATTACGGCTGTCAATGCGGAACGGTTTCAGGACCCCGCTGACCTGGCGCGCAAGGTT 1399
QY 1093 ggcgcattacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1152
Db 1400 GCCAATATCGCCCTGTGAAAAGGCGGCCCTGACCGCTCTGGCGCAAGAACAAAGCGCGAG 1459
QY 1153 acaatcaagcgaagctgg 1171
Db 1460 GAATCAACGTCACGATTG 1478

RESULT 6
BRUHTRAH
LOCUS BRUHTRAH 1905 bp DNA BCT 28-NOV-1994

```



```

/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/clone_lib="Kohara phage"
56..1577
/gene="degQ"
56..61
/gene="degQ"
79..84
/gene="degQ"
91
/misc_feature
/feature="degQ"
/feature="primer extension identifies this as the
transcription start site (+1 position)"
126..138
/gene="degQ"
136..216
/gene="degQ"
/evidence=experimental
136..1503
/gene="degQ"
/feature="serine protease in vitro"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DegQ"
/protein_id="AAC44005.1"
/db_xref="GI:984379"
/translation="MKQTQLLSALSVGLTISASFOAVASIPGOVADAPLPISLAP
MLKVLPAVVSVEGTASOGKIPFEKFFGDDLPDPAQPEFGSGVLIINASKG
VYLTNNHVINQAQKISIQLNDRREFDAKLIGSDQSDIALLOIONPKLQIAIADSD
KLVRGFAVAGFPFGQATSGIVSALGRSLNLEGFNIOTDASINRNGSGGAL
LNNGELIGINTAILAPGGSVGIGFAIPSNMARTLAQQLIDFGEIRGLGIRGTEM
SADIAKAFNDVQRFVSVLPFGSSAKAGVAGDIITSLNGKPLNSFABELRSRIAT
TEPTKVKLGLLNKRPLEVEVTLDTSTSSASAEITTPALEGATLSDGLKDGKGI
KIDVVKGSPAAQGLQKDDVIGVNRDRVNSIAEMRKVLAAPAILALQIVRGNESI
YLLMR"
217..1500
/gene="degQ"
/feature="serine protease in vitro"
/evidence=experimental
/product="DegQ"
1508..1547
/gene="degQ"
1508..1577
/gene="degQ"
/feature="consistent with degQ transcript size defined by
Northern analysis; Rho independent: stem loops followed by
a series of t residues"
1532..1537
/gene="degQ"
1532..2709
/gene="degQ"
1550..1571
/gene="degQ"
1554..1559
/gene="degQ"
1566
/gene="degQ"
/feature="primer extension identifies this as the
transcription start site (+1 position)"
1579..1595
/gene="degQ"
1593..2660
/gene="degQ"
/feature="putative serine protease"
/codon_start=1
/transl_table=11
/product="DegS"
/protein_id="AAC44006.1"
/db_xref="GI:984380"
/translation="MFVKLLRSVAIGLVGAILLVAMPFSRLSNPLSTPOFDSTDETP
ASYNLAVRRAPAVVNVNRLNTNSHNQLEIRTLGSGVIMDQRGYIITNKHVINDAD

```

```

OLIVALQDGRVFEALLVGSDSLTDLAVLKINATGGLTIPINARRYPHICDVVLAIGN
PINLQITITQGIISATGRIGNPTGRNQFLQTDASINHGNSGALVNSLGMINTL
SFDKSNVGETPEGIFAIPFQALIMDKLIRDRVIRGYIGIGIAPLPHAOGGI
DQLAQIVVNEVSPDGPAAQVNDLIISVDNKPASIALETMDQVAEIRPGSVIPV
VMRDKQLTQVITQIEYPATN"
2674..2709
/feature="degS"
2674..2709
/feature="degS"
/feature="consistent with degS transcript size defined by
Northern analysis; Rho independent: stem loop containing a
series of t residues"
659 a 692 c 710 g 653 t
ORIGIN

Query Match 14.1%; Score 196; DB 2: Length 2714;
Best Local Similarity 53.6%; Pred No. 1.7e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 382 ggcttcacatcagcaaaacggctacatcctgaccacaataccaccgctgtgcccgtatg 441
Db 418 gTCATCATCAACGCCAGTAAAGGCTATGTCTGACCAACAACCATGTGATTAATCAGGCA 477
QY 442 ggcagtatcaaaagtcctgctcaacgacgacgaataataccgcaaaactcaggttcg 501
Db 478 CAGAAAAATCAGTATTTCAGTCAATGATGGCGCGAGTGTGATGATAAAACATGTTGGTAGC 537
QY 502 gattgccatcagatgctccctctgaaatcgacgcaacggaagagctaccgcgtcgc 561
Db 538 GATGACGAGGAGATATCCCTGTGTACAAATCAAAACCCGAGCAAAATTAACGCAAAATC 597
QY 562 aaaaatcggcaatccccaaaattgaaacccggcggaatgggtcgtccatcgcgcgcgcc 621
Db 598 GCTATTGCCGACTCCGATAAATGCCGCTCGGTGATTTTCCCGTAGCGGTCGGTAACCCA 657
QY 622 ttccgctttgacaacagcgtgaccgcggcatcgtgctccgcaaaagcagaagccctgcgc 681
Db 658 TTTGGCCTTTGGCAAAACCCGCCACCTCTGGCATTTGTTCCGCATTTAGGCCGACGCGGTTG 717
QY 682 aa---cgaaagctacacaccccttcacaaacgacgcttgccatcaatcgcggcaattcc 738
Db 718 AATCTTGAAGGTCGTGAAAAACATTTATCCAGACAGATGCTTCCATTAACCCGCGTAACCTC 777
QY 739 ggcggccgcgtgttcaacttaaaaggacaggtcgtcggcatcaattcgcacaataatacagc 798
Db 778 GCGCGTGCACTATTAAACCTTAACGGTGAGTTAATTTGGCATCAACACTGCAATCCTTTGCG 837
QY 799 cgcagcggcggtattcatgggcatctccttggccatcccgatgacgttgccatgaatgctc 858
Db 838 CCTGGCGCGGAGCGTCGGGATTGGATTGGCATCCCGAGTAATATGGCGCGAAACACTG 897
QY 859 gccgaacagctgaaaaacacccggcaaaagtcacacgcggaacactggcggtgattattcag 918
Db 898 GCGCAGCAGCTTATCCACTTTGGTGAATCAAAACGGGTTTGTGTAGGCATCAAAAGGCACC 957
QY 919 gaagtatctacaggtttggcacagtcgttcggtctctggataaaagccagcgcgcatgatt 978
Db 958 GAGATGAGTGCGGATATGCCAAAGCCTTCAACCTTTGACGTGACGCGTGCGCGGTTTGTC 1017
QY 979 gccaaaatccttcccggaagccccgcagaacgtgcccgtgcagcggcgagacatgctc 1038
Db 1018 ACAGGAAGTGTTCGCGAGTTCTTGGCTCGGCAAAAGCGGGCGGTCAAGCGCGGATATTATT 1077
QY 1039 ctccagcctcagcgggagagaataacttctccggcgacacttccctcactggtcggcgcc 1098
Db 1078 ACCAGCCTCAACGGCAAAACCGCTGAATAGCTTTGCTGAGTTGCGCTCTCGTATCGCGACC 1137
QY 1099 attacgcggggaaaaagaagtcagcctcgcgtatgctgcgaagggcgagaataatacaatc 1158
Db 1138 ACCGAGCCGGGACCAAAAGTGAAGCTTTGGCTGCTGCTGCGTAACGCCAAACCACTGGAAGTA 1197
QY 1159 aaagccaagctgggcaacgcccgc 1182

```



```

Db 1198 GAAGTGAGGCTCGATACCAACACC 1221
|||||
RESULT 8
ECU15661 3156 bp DNA BCT 25-APR-1996
LOCUS
DEFINITION Escherichia coli serine protease (hhoA and hhoB) genes, complete
cds, and malate dehydrogenase (mdh) gene, partial cds.
ACCESSION U15661
VERSION U15661.1 GI:558911
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 3156)
Bass,S., Gu,Q. and Christen,A.
Multicopy suppressors of prc mutant Escherichia coli include two
HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated RipA
J. Bacteriol. 178 (4), 1154-1161 (1996)
MEDLINE 96165273
REFERENCE 2 (bases 1 to 3156)
Bass,S.
Direct Submission
Submitted (07-OCT-1994) Steve Bass, Cell Genetics, Genentech Inc.,
460 Point San Bruno Blvd, South San Francisco, CA 94080-4990, USA
FEATURES
source
1..3156
/organism="Escherichia coli"
/strain="W3110"
/db_xref="taxon:562"
/map="70 minutes"
454..459
-35_signal
477..482
-10_signal
534..614
sig_peptide
/gene="hhoA"
534..1901
gene
/gene="hhoA"
534..1901
CDS
/gene="hhoA"
/feature="putative serine protease"
/codon_start=1
/transl_table=11
/product="hhoA"
/protein_id="AAC43992.1"
/db_xref="GI:558912"
/translation="MKQTQLLSALSVGLTSLASTCAVASIPGVADQAPLPSLAP
MLEKLVPAVSVRVEGTASQKQIPEEFKFFGDDLPDQPAQPEGLSGVINASKG
YVLTNNHVIHQAKISIQINDREFDAKLIGSDSDIALLIQNPISKLQIAIADSD
KLRYGDFAVAVNPFGLGQTATSGIVSALGRSLNLEFNFIQTASINRNGSGAL
LNLNGELIGINTAILAPGSGVGIGFAIPSNMARTLAQLIDFGEIKRGLIGKTEM
SADTAKFNLDVQRAVSEVLPVSGSAKAGVAGADITISLNGKPLNSFABLRRIAT
TEPTKVKLGLLRNKPVEVETLDTSTSSASAEIMITPALEGATLSDGOLKDGKGI
KIDEVVGSPPAAQGLQKDDVLIIGVNRDRVNSIAEMRKVLAAPAILALQIVRNEI
YLLMR"
615..1898
mat_peptide
/gene="hhoA"
/product="hhoA"
1906..1945
stem_loop
1991..3058
gene
/gene="hhoB"
1991..3058
CDS
/gene="hhoB"
/feature="putative serine protease"
/codon_start=1
/transl_table=11
/product="hhoB"
/protein_id="AAC43993.1"
/db_xref="GI:558913"
/translation="MFVKLLRSVAIGLVGAILLVAMPSLRSLNPLSTPQFDSTDETP
ASYNLAIVRAAPAVVNVYINRGLNTNSHNOLEIRTLGSGVIMDQRGYIITNKHVINAD

```

```

QIIVALQDRVFALLVGSDSLTDLAVLKINATGTLPTIPINARRVPHIGDVVLAIGN
PYNLGOTITQGLISATIGRLNPTQRNQLFQDASINHGNSGALVNSIGELMGINTL
SFDKNDGETPEGIFAIPFQIATKIMDKLIRDGRVIRGVIIGIGREIAPLHAQGGI
DOLOGIVVNEVSPDGPAAAGIQTVDLIISVDNKPISALETMDQVAEIRPGSVIPV
VMRDKQLTLQVTIQEYPATN"
3070..3109
terminator
/feature="potential terminator for hhoB and mdh genes"
complement(3117..3156)
gene
/feature="mdh"
complement(3117..3156)
CDS
/feature="mdh"
/codon_start=2
/transl_table=11
/product="malate dehydrogenase"
/protein_id="AAC43994.1"
/db_xref="GI:558914"
/translation="ISPWAKSSLIISN"
BASE COUNT 768 a 807 c 828 g 753 t
ORIGIN
Query Match 14.1%; Score 196; DB 2; Length 3156;
Best Local Similarity 53.6%; Pred. No. 1.7e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;
QY 382 ggcttcacatcagcaaaacggctacatcctgacccaataccacgctcgtgcccgtatg 441
Db 816 GTCATCATCAACGCCCAAGGCTATGCTGCTACCAACACCATGATGATTAACAGGCA 875
QY 442 ggcagtatcaaatcctgctcaacgcaacgcaataataccgcaaaactcatcggttcg 501
Db 876 CAGAAATCAGATTCAGCTCAATGATGGCGCGAGTTGATGCAAACTGATTGGTAGC 935
QY 502 gatgtccaatccgatgtgcocctctctgaaatacgacacgcaagagctaccgctcgc 561
Db 936 GATGACGACGAGGATATCGCCTGTTACAAATTCAAAACCCGAGCAATTAACGCAATC 995
QY 562 aaatcggaatcccaaaatttgaaacccggcgcaatgggctgctgcccactcgcgccc 621
Db 996 GCTATTGCCGACTCCGATAAATTCGGCTCGGTGATTTGCGGTAGCGGTACGTAACCA 1055
QY 622 ttccgctttgacacagcgtgacccggtgcatcgtgtccccaagcgagaagcgtgcgc 681
Db 1056 TTTGGCCTTGGCAACCGCCACCTCTGCGATTGTTCCGCATTAGCGCGAGCGGGTTG 1115
QY 682 aa---cgaaagctacacaccccttcacaaaccacgactgctgcataatcccggaattcc 738
Db 1116 AATCTTGAAGGCTCTGGAATACTTTATCCAGACAGATGCTTCCATTAAACCGGTAACTCC 1175
QY 739 ggcggccgctgttcaacttaaaaggacagctgctcgcatcaatcgcgaatatcacagc 798
Db 1176 GCGGTGCACATATTAAACCTTAAAGTGAGTAAATTGCGATCAACACACACACACACACAC 1235
QY 799 cgcagcggcgattcatgggcatctcctttgccatcccgattgacgttgccatgaatgc 858
Db 1236 CCTGCGCGGGAGCGGTGCGGATTGGATTGGCATCCCGAGTAATATGCGCGGAACACATG 1295
QY 859 gccgaacagctgaaaaacacccggcaaaagtcacacccgacacactggcggtattattcag 918
Db 1296 GCGCAGCAGCTTATCGACTTTGGTGAATAACAACCGCGTTTGTAGGCATCAAGGCGAC 1355
QY 919 gaagtatcctacggtttggcacagctcgttcgtctggataaaagccagcgccattgatt 978
Db 1356 GAGATGAGTGGCGATATCGCCAAAGCCCTTCAACCTTGACGTGACGCGTGGCGCGTTTGT 1415
QY 979 gcaaaatccttcccgcgagcccccgcaaacgctgcggtcagcgcgagcgacatcgtc 1038
Db 1416 ACGGAGGTGTTGCCAGGTCTTGGCTCGGCAAAAGCGGGGTCAAAAGCGCGCATATTATT 1475
QY 1039 ctacagctcagcggcgagaaaaatcgttttccggcgacccctcccgctacatgctcgcgcc 1098
Db 1476 ACCAGCCTCAACGCGCAAAACCGCTGAATAGCTTTGCTGAGTTGCTGCTCTCGTATCGGACC 1535

```



```
/db_xref="GI:1789622"
/db_xref="PID:g1789622"
/translation="MEGLIISVILITWGYLLKNYKPOVLAAGIFLMGCVWLGF
CGVDPSTSSGLYVDINEILRMSNRAGLSIMAVGGTARTMERIGASRAMVSL
LSRPUKLRSPYIILUSATYIGQIAQFITSASGLGMLMLVTLFPTLVLSGVSRLSA
AVIATITSEINGLITNSIFAQVAGMKIATFFHYQLPVSACVITSVAISHFFVORA
FDKKDNIHQEAKQALDNVPLYYAILPVMPLILMLGSLFLAHVGLMQSELHVVV
MLLSLTVMEVFFPKHNLRETMDVDQAFEDGMGTQFANVTVLVVAGETFAKGLTTIG
TVDAVIRGAHSGLGIGVIMIALVIALCAIVMSGNAPFMSFASLIINIAAGLHP
AVVIMPHFATTLARASPIITAVVVVTSIAGVSPFAVVKRTAIPMAVGFVYVNIAT
ITLF"
complement(3666..4163)
/gene="sspB"
/gene="b3228"
complement(3666..4163)
/gene="sspB"
complement(3666..4163)
/function="regulator; Global regulatory functions"
/note="f165; 100 pct identical amino acid sequence and
equal length to SSPB_ECOLI SW: P25663"
/codon_start=1
/transl_table=11
/product="stringent starvation protein B"
/protein_id="AAC76260.1"
/db_xref="GI:1789623"
/db_xref="PID:g1789623"
/translation="MDLSQLTFRPRPYLLRAFYEWLLDNQITPHLYVDVTLPGVQVPM
YARGQIVLNTAPRAVGNLEANDVRFNAREGGIPQVSVPLAAVLAIAIARENGAGT
MFEPFAAYDEDTISINDEASADNETVMSVIDGDKPDHDDTHPDDPEPPPPRGRPA
LRVK"
complement(4166..4194)
/note="factor Sigma70; predicted +1 start at 3374407"
complement(4169..4807)
/gene="sspA"
complement(4169..4807)
/note="b3229"
complement(4169..4807)
/gene="sspA"
/function="regulator; Global regulatory functions"
/note="f212; 100 pct identical amino acid sequence and
equal length to SSPA_ECOLI SW: P05838"
/codon_start=1
/transl_table=11
/product="regulator of transcription; stringent starvation
protein A"
/protein_id="AAC76261.1"
/db_xref="GI:1789624"
/db_xref="PID:g1789624"
/translation="MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSEIEHVEKD
NPQDLIDLPNQSVPTLVDRELTWESRIIMEYLDERPHPLMPVYPVARGESRLY
MHRTEKDYTLWNTIINGSASADAARKOLRELLAIAVFGOKPVFLSDPSLVDCY
LAPLLWRPLQGLIESGPGAKELKGYMTRVFRDSEFLASLTAEEREMRLGRS"
complement(4944..4972)
/note="No predicted promoter"
/bound_moiety="Galn predicted site"
complement(5202..5594)
/gene="rpsI"
/note="b3230"
complement(5202..5594)
/gene="rpsI"
/function="structural component; Ribosomal proteins -
synthesis, modification"
/note="f130; CG Site No. 17596; 100 pct identical
amino acid sequence and equal length to RS9_ECOLI SW:
P02363"
/codon_start=1
/transl_table=11
```

```
Query Match 14.1%; Score 196; DB 2; Length 10713;
Best Local Similarity 53.6%; Pred. No. 1.6e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 382 ggcttcacacgcaaaacgctacatctgaccaataaccacgctgtgctgcggtatg 441
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8412 GTCATCATCAAGCCAGTAAGGCTATGTCTGACCAACACCAACCATGTGATTATCAGGCA 8471
```

```
Qy 442 ggcagtatcaaaagtcctgctcaacgacagcgaataatccgccaaactcactcggttcg 501
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8472 CAGAAATACGATATTCAGCTCAATGATGGCGCGAGTTTGATGCAAAACTGATGTAGC 8531
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 502 gatgccaaatccgatgtcgccctctgaaatcgacgcaacggaagagctacccgcgtc 561
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8532 GATCAACAGAGCATATCGCCCTGTTACAAATTCAAACCCGAGCAAAATTACGCCAAATC 8591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 562 aaatcggaatccccaaaaatttgaacccggcggaattgggtcgctgceatcgcgcc 621
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8592 GCTATTGCCGACTCCGATATAATTGCGCTCGTGCTGATTTGCGGTAGCGTCGTAACCCA 8651
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 622 ttggctttgacaacagcgtgacccggtcgatcggtgcgtgcgaagagcagaagcctgc 681
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8652 TTGGCTCTGGGCAACCGCCACCTCTGCGATTTGTTCCGATTTAGCGCGCAGGGGTG 8711
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 682 aa----gaaagctacacaccccttcataccaaacggcggttgccatcaatccgggcaatcc 738
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8712 AATCTGAAGGCTCGGAAACTTTATCCAGACAGATGCTTCCATTACCGCGGTAACTCC 8771
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 739 ggcggccgcgtgttcaactaaaaggacaggtcgctcggtcgcataatccgaataatcacgc 798
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8772 GCGGTGCACTATTAAACCTTAACGGTGAGTTAATTGGCATCAACACTGCAATCCTTGGC 8831
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 799 cgaagcggcggttcattcgtggcatctcctttgccatcccgattgacgttgccatgaatgc 858
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8832 CCTGCGCGCGGAGCGTGGGATTGGATTTGCCATFCCCACTCAATATGCGCGAACAACACTG 8891
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 859 gccgaacagctgaaaaaacacccgcaaaqtcacacgcgcgaactggcggtgattattcag 918
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8892 GCGCAGCAGCTTATCGACTTTTGGTGAAATCAACCGCGTTGTTAGGCATCAAGGCACC 8951
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 919 gaagtactctacggtttggcacagctcgttcggtctggataaaagccagcgcgcatgtgatt 978
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8952 GAGATGAGTGCAGATATCGCCAAAGCCCTTCAACCTTGACGTGCAGCTGGCGGCTTTGTC 9011
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 979 gccaaaatccttcocggcagccccgcgagaaagtcgcccgcctgcagggcgacatcgtc 1038
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9012 AGCGAAGTGTTCAGAGTTCTGGCTCGCNAAGCGCGCTCAAAAGCGGCGATATTATT 9071
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1039 ctgagctcgacgcgagaaataacgtcttcctccgcgacctcccgctcatgctcgcgcc 1098
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9072 ACCAGCTCAACGCCAAACCCCTGAATAGCTTTGCTGAGTTGCGGTCTCTGATCGGAC 9131
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1099 attacgcggaagaaagtagtcagctcctggcgtatggcgaaggaagaaatacaatc 1158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9132 ACCGAGCGCGGCACAAAGTGAAGCTTGCGCTGCTGCTAACGGCAAAACCACTGGAAGTA 9191
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1159 aaagccaaagctgggcaacgcgcgc 1182
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9192 GAAGTGACGCTCGATACCAGCAC 9215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
ECOUW67_1
WPCOMMENT
Sequence split into 4 fragments LOCUS ECOUW67 Accession U18997
Fragment Name Begin End
ECOUW67_0 1 110000
ECOUW67_1 100001 210000
ECOUW67_2 200001 310000
ECOUW67_3 300001 372438
Continuation (2 of 4) of ECOUW67 from base 100001 (U18997 Escherichia coli K-12 chro
```

```
Query Match 14.1%; Score 196; DB 2; Length 110000;
Best Local Similarity 53.6%; Pred. No. 1.6e-30;
Matches 431; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 382 ggcttcacacgcaaaacgctacatctgaccaataaccacgctgtgctgcggtatg 441
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61770 GTCATCATCAAGCCAGTAAGGCTATGTCTGACCAACACCAACCATGTGATTATCAGGCA 61829
```

Qy	442	ggcagtatcaaaagtctctgctcaagacagcgcaataaccgccaactcatcggttcg	501
Db	61830	CAGAAAAATCAGATTACGCTCAATGATGGCGCGAGTTTGATGCAAAACTGATTGGTAGC	61899
Qy	502	gatgtccaaatccgatgtgcgcctctgaaatcgcagcaaggaagagctaccgcgcgtc	561
Db	61890	GATGACCAAGACGATATCGCCTGTACAAATTCAAAAACCGCAGCAAAATTAACGCAAAATC	61949
Qy	562	aaatcgcgcgaatcccaaaaattgaaaccggcgcaatggctgcgtgcgcacatcgcgcgcc	621
Db	61950	GCTATTGCCGACTCGGATAAATTGCGCGTCGGTGATTTTGGCGTAGCGGTGCGTAACCCA	62099
Qy	622	ttcggctttgacaacagcgtgaccgcgcgcgtctgctccgcgaaggcagaagcctgcgc	681
Db	62010	TTTGCGCTTGGCAAAACCGCCACCTCTGGCATTTGTTCCCGCATTAAGCGCGCGGGTTG	62069
Qy	682	aa---cgaagactcacacaccttcatcaaacaccgaagctgtgcatacaatcccggaattcc	738
Db	62070	AATCTTGAAGTCTGGAAACCTTTATCCAGACAGATGCTTCATTAAACGGGGTAACATCC	62129
Qy	739	ggcgccgcgtgttcaacttaaaaggacagctgcgcgcacataattcgcaaatatacagc	798
Db	62130	GGCGGTGCACATTAAACCTTAAGCGTGAGTTAATTGGCATCAACACTGCAATCCTTGCG	62189
Qy	799	cgcagcgcgcgtattcgggcgtatcctttgccatccgcgttgcctgcatgaatgct	858
Db	62190	CCTGGCGCGGAGCGTCGGGATTGGATTGCCATCCCGAGTAATATGGCGCAACACTG	62249
Qy	859	gccgaacagctgaaaaacaccgcgcaaaagtcacacgcgcgaacactggcgctgatttcag	918
Db	62250	CGCAGCAGCATTATCGACTTTGGTGAATCAAAACCGCGTTTGTAGGCATCAAAAGGCACC	62309
Qy	919	gaagtatcctacgctttggcagcagctcgcttcggtctggataaagccagcgcgcattgatt	978
Db	62310	GAGATGAGTCCGATATCGCCAAAGCCCTTAACCTTGACGTGCACGCTGGCGCGTTGTC	62369
Qy	979	gccaaaaatccttcccggcagcccgcgacagacgctgcgcgcctgcagcgcgcgacatcgtc	1038
Db	62370	AGCGAAGTGTTCGCCAGTCTTGCGTCGGCAAAAGCGGCGCTCAAAAGCGGCGATATTATT	62429
Qy	1039	ctcagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1098
Db	62430	ACCAGCCTCAACGGCAAAACCGCTGAATAGCTTTGCTGAGTTGCGCTCTCGTATCGCGACC	62489
Qy	1099	attacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1158
Db	62490	ACCGAGCGCGGCACGAAGTGAAGCTTGCGCTGCTGCGTGAACGGCAAAACCACTGGAAGTA	62549
Qy	1159	aaagccaagctgggcgaacgcgcgc	1182
Db	62550	GAAGTGACGCTCGATACCAAGCACC	62573
RESULT	11		
PAU29172			
LOCUS	PAU29172	1191 bp	DNA
DEFINITION	Pseudomonas aeruginosa HraA-like serine protease	BCT	06-MAR-1996
	complete cds.		
ACCESSION	U29172		
VERSION	U29172.1	GI:1184674	
KEYWORDS			
SOURCE			
ORGANISM			
	Pseudomonas aeruginosa.		
	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;		
AUTHORS	Pseudomonas.		
	1 (bases 1 to 1191)		
TITLE	Boucher,J.C., Martinez-Salazar,J., Schurr,M.J., Mudd,M.H., Yu,H.,		
	and Deretic,V.		
	Two distinct loci affecting conversion to mucoidy in Pseudomonas		
	aeruginosa in cystic fibrosis encode homologs of the serine		
JOURNAL	protease HraA		
	J. Bacteriol. 178 (2), 511-523 (1996)		

MEDLINE	96134987
REFERENCE	2 (bases 1 to 1191)
AUTHORS	Boucher J.C.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUN-1995) John C. Boucher, Microbiology, UTHSCSA, 7703 Floyd Curl Dr., San Antonio, TX 78284, USA
FEATURES	Location/Qualifiers
source	1..1191
	/organism="Pseudomonas aeruginosa"
	/strain="PAOI"
	/db_xref="taxon:287"
CDS	/map="69 min"
	14..1183
	/note="HtrA-like serine protease"
	/codon_start=1
	/transl_table=11
	/product="AlgW"
	/protein_id="AAC43674.1"
	/db_xref="GI:1184675"
	/translation="MPKALRFGLGWPVLGVLLALLIQHNPLVGLPROEVHVHEQAPL LSRLQEGSYANAYASRAAPAVANLVTKMKRPSHPFLDDPMFRFFGDNLPOOKRM ESLSGSAYMSAGEYILLTNNHTAGADQIIVALRDGRETI AOLVGSDPETDLAVLID LNKLPMATLRSGDITGTDCVLAINGFPGVGGTRDRGHGRHRHQTPAREDTTDFLOT DAINLPNGSGVALDAAGNLTGINTAIKFSGSGQSIRSPSRPTVMYSQTIEHGQ VI RWTGVGEVKALTPELASLGLETIGIVAGRLSGRAPAGGLGPGLDVILTIDROE ASDGRSMNQVARTRGQIKISIVLRNGKVNLTAEBVLGRPAPPAPQQKDQG"
BASE COUNT	214 a 428 c 376 g 173 t
ORIGIN	
Query Match	13.4%; Score 186.6; DB 2; Length 1191;
Best Local Similarity	54.4%; Pred. No. 1.4e-28;
Matches	462; Conservative 0; Mismatches 379; Indels 8; Gaps 4;
Qy	367 ttgaacttcggttgccgcctcattcaatcgacgaaacaagcggtacctctgcaccataccccac 426
Db	
	323 TCGAGCCTCGCCTCGGCCTCATGACGCCGAAGGCTACTCTGTGACCACAACCCAC 382
Qy	427 gtggttgccgtatggcgagtcaaaaagtcctgctcaacgagaagaagggaatataccgcc 486
Db	
	383 GTGACCCTCGGCCCGCACCAAGTCATCGTGCCCTTCGCCGACGCCGCGAAACCATCGCC 442
Qy	487 aaactcatcgttgcgatgccaatccgatctgcctctctgccttctaaaaatcgacgcaaggaa 546
Db	
	443 CAGTTGTCGGCAGCGACCCGGAAACCGACCTGGCGCTGTGAAGATCGACCTTA --- AG 499
Qy	547 gagctaccctcgtaaaatcgcgaatcccataaaattgaaacccggcggaatggctcgt 606
Db	
	500 AACCTGCGCGATGACCCCTCGCGCTCGCGAGGATTCGCACCGCGACGCTGCGCTC 559
Qy	607 gcatacgcgcccttcggcttgtgacaaa -gcgtgaccgcgggatactggttcgcgcaa 665
Db	
	560 GCCATCGGCAACCCGTTTCGGCTCGGCCACACCCTGACCATGGCATATCAGCGCAC 619
Qy	666 aggcagaagcctgcaccaac--gaaagctcacacccttcacaaaccgacgttgccatc 723
Db	
	620 CGNCGCAACAGCTCGGCTGTACCTACAGAGATTATCATAGACCGAGCGGCGATC 679
Qy	724 aatccgggcaattccggcgcccgctgttcaacttaaaggacaggyctgcggcatacaat 783
Db	
	680 AACCCTGGCAACTCGGCGCGCGCTGGTGGACGCTGCCGGCAACCTGATCGGCATCAAC 739
Qy	784 tcgcaaatatacagccgcgaagcgcgatcatgggcatctcctttgccatcccgattgac 843
Db	
	740 ACGGCGATCTTCCAAGTCCGGCGCTCCAGGGTATCCGTTTCGCCATCCCCACCAAGC 799
Qy	844 gtfgcatgaatgtcgcgcaacagctgaaaacacacccgcgaaagtccaacgcgcaacaactg 903
Db	
	800 TGNACCTGGAGTATGCAGTGCATCATGCACAGCGGCAGGTGATCCCGGCGTGCTC 859
Qy	904 ggcgtagattacggaagtattactacggtttggcacagtcgcttgcgctcggtataagcc 963
Db	
	860 GGCTCGAGCTCAAAGCGCTGACCCCGGAATCTGGCGAGTCTGCTGGGCTTCGCGAAACC 919


```
ODLLHRVHGNIKWLGMGAIGGLGCLLGFVAPATSGGFGNLIPIATAGNFSGMLV
FIFVARVITLLCFSSGAPGIFAPMLAUGTVLGTAFGMVAVELFPDYLHLEAGTFAIA
GMLALAAASIRAPLTGILVLEMTDNYQLIPLMIITGLGATLQAFTGGKPLYSAILA
RTIAKQAEQALARSKAASASENT"
3031. 3059
/notes="factor Sigma70; predicted +1 start at 176559"
3052. 3082
/notes="factor Sigma70; predicted +1 start at 176582"
3082. 3114
/notes="factor Sigma70; predicted +1 start at 176614"
3117. 3461
/genes="yadR"
/notes="b0156"
3117. 3461
/genes="yadR"
/function="orf; Unknown"
/notes="o114; 100 pct identical to YADR_ECOLI SW: P37026"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73267.1"
/db_xref="GI:1786351"
/db_xref="PID:g1786351"
/translation="MSDDVALPLEFTDAANKVKLSIADENPNLKLRVYITGGCSG
FOYGFEDDQVNEGDMTIEKQGVGLVDPMSLQYLVGGSYDYTEGLEGSRFIYTPNA
KSTCGGGSFSI"
complement(3508. .4131)
/genes="yadS"
/notes="b0157"
complement(3508. .4131)
/genes="yadS"
/function="orf; Unknown"
/notes="f207; 100 pct identical to YADS_ECOLI SW: P37027"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73268.1"
/db_xref="GI:1786352"
/db_xref="PID:g1786352"
/translation="MLVYWLIDVGTAVFAISGVLLAGLRMDPFGVLVLGVTVAGGG
TTRDMALDGPVFWKDPDTDLVVMVTSMLTIVLQRPRLPKWMLPVLDAVGLAVFV
GICVKNFAEAGPLIACVGMVITGVGGGIIRDVLAREIPMLIRTEIYATACIIGIV
HATAYTFVSVPLETSAMGMGMVTVLLIRLAIRHLKPLTALDENG"
complement(4169. .4969)
/genes="yadT"
/notes="b0158"
complement(4169. .4969)
/genes="yadT"
/function="orf; Unknown"
/notes="f266; 99 pct identical to YADT_ECOLI SW: P37028"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73269.1"
/db_xref="GI:1786353"
/db_xref="PID:g1786353"
/translation="MAKSLFRVALVSFLAPLMLNAPRVITLSPANTELAFAAGITP
GVSSYSDTPPQAQKEQVSTWQGMNLERIVALKPDLVIAWRGNGNAERQVDLASLGI
KVMVDATSIETANALRLAPSPQPDRAEQAAQSLDQYLAQKAYADKPKRRVFL
QFGINPFTSGKESIQNVLEVGGENIFKDSRPVPOVSRQVLRASPOAIVITGSP
DOIPTKQYWGEOLEKIPVLPVLTSDWFERASPRILAAQQLCNALSQVD"
complement(4212. .4239)
/genes="yadT"
/notes="factor Sigma70; predicted +1 start at 177698"
complement(4962. .5660)
/genes="pfs"
/notes="b0159"
complement(4962. .5660)
/genes="pfs"
/function="orf; Unknown"
/notes="f232; 100 pct identical to PFS_ECOLI SW:
P24247; purine nucleoside phosphorylase (deob) similarity"
/codon_start=1
```

```
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73270.1"
/db_xref="GI:1786354"
/db_xref="PID:g1786354"
Query Match 12.8%; Score 178; DB 2; Length 10199;
Best Local Similarity 52.1%; Pred. No. 7.5e-27;
Matches 422; Conservative 0; Mismatches 385; Indels 3; Gaps 1;
Qy 382 ggcttcacatcagcaaaaacgctacatccctgacacaataccacacgctcgttcggtatg 441
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7739 GTCATCTATTGATGCCGATTAAGGCTATGTGTCACCAACAACACCGTGTGTGATACGGC 7798
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 442 ggcatatcaaaagtctctcaacgacagcgcaaatataccgccaactcatcgggttcg 501
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7799 ACGGTCAATTAAGATTCAACTGAGCGATGGCCGTAAAGTTCGACGCGAAGATGGTTGCCAAA 7858
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 502 gatgtccaatccgagtgcgcctctctgaaatctcgacgcaacggaagagctaccctcgtc 561
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7859 GATCCGCGCTCTGATATCGCGTGTATCCAAATCCAGAACCCGAAAAACCTGACCGCAATT 7918
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 562 aaaaatcgcaatcccaaaaatttgaacccggcgcaatgggtcgctgccatcgcgcccc 621
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7919 AAGATGGCGGATTCGTATGCATCGCGTGGTGGTGTATACACCGTAGCGATTGTTAACC 7978
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 622 ttgggtcttgacacagcggtgaccgccgcatcgctgcgcgcaaaag---gcagaagccctg 678
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7979 TTTGGTCTGGGCGAGACGCGTAACCTCCGGGATTGTCTCTGCTGGGCGGTAGCGCCCTG 8038
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 679 cccaacgaaagtacacaccccttcacaaacgacgctgtgcatcaatccgggcaattcc 738
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8039 AATCGCGAAACTACGAAACATTCATCCAGACCGATGACAGCATCAACCGTGGTAACTCC 8098
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 739 ggcgcgcgcgttcaacttaaaaggacaggtcgctgcgcatcaatcgcaaatatcacagc 798
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8099 GGTGGTGGCTGGTTAACTGAACGGCGNACTGATCGGTATCAACACCGCGATCTTCGCA 8158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 799 gcagcgcgcgattcattgggcttcctttgccatcccgatgacgttgcattgaatgct 858
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8159 CCGGACGGCGGCAACATCGGTATCGGTTTGTCTATCCCGAGTAACATGGTGAACAAACCTG 8218
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 859 gccgaacagctgaaaaacacacgcgcaaaagtccaacgcggacacactggcggtgattatcag 918
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8219 ACCTCGCAGATGGTGGAAATACGGCCAGGTGAACACCGGTGAGCTGGGTATTATGGG 8278
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 919 gaagtatcctacggtttggcacagtcgttcggtctgtgataaaagccagcgcgcatgatt 978
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8279 GAGCTGAACCTCCGAACTGGCGAAACGGATGAAAGTTGACGCCCGCCGCGGTGTTTCGTA 8338
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 979 gccaaaaatccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1038
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8339 AGCCAGGTTCTGCCCTAATTCCTCCGCTGCAAAAAGCGGGCATTAAGCGGGTGTATGTATC 8398
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1039 ctacgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1098
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8399 ACCTCACTGAACGGTGAACCGCATACAGACGCTTTGCCGCACTGCTGCTCAGTGGGTACT 8458
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1099 attacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8459 ATGCCGCTAGGCAGCAAACTGACCTTGCGCTTACCTGCGGACGGTAAGCAGGTTAAGCTG 8518
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1159 aaagccaaagctgggcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1188
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8519 AACCTGGAACCTGCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 8548
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 13
ECOB2K 82727 bp DNA BCT 12-SEP-1997
LOCUS Escherichia coli genome, 2.4-4.1 min region (110,917-193,643 bp
DEFINITION from 0 min).
ACCESSION D26562
```

- VERSION D26562.1 GI:473770
KEYWORDS 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase; GMP reductase; PII interlyl-transferase; adenosylmethionine decarboxylase; proenzyme; ampd signalling protein; ampe signalling protein; aromatic amino acid transport protein; deoxyguanosine triphosphate triphosphohydrolase; dihydroliipoamide acetyltransferase; dihydroliipoamide dehydrogenase; dosage-dependent dnaK suppressor protein; elongation factor EF-rs; ferrichrome-iron receptor; ferrichrome-iron transport protein fhuB; ferrichrome-iron transport protein fhuC; ferrichrome-iron transport protein fhuD; glucose dehydrogenase; glutamate-1-semialdehyde aminotransferase; heat shock protein htrA; methionine aminopeptidase; mukB suppressor protein snbA; mutator mutT; nicotinate-nucleotide pyrophosphorylase (carboxylating); penicillin-binding protein 1B; plasmid copy number control protein pcnB; pyruvate dehydrogenase; regulatory protein fsl1; ribosomal protein S2; ribosome releasing factor; spermidine synthase; tetrahydridipicolinate N-succinyltransferase; Escherichia coli (sub_strain W3110, strain K-12) (library: Kohara's mini set library) DNA, clone #111, 112, 113, 114, 115, 116 and 119. Escherichia coli
- SOURCE Escherichia coli
- ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
- REFERENCE 1 (bases 1 to 82727)
AUTHORS Fujita, N.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-1994) to the DDBJ/EMBL/GenBank databases.
AUTHORS Nobuyuki Fujita, National Institute of Genetics, Department of Molecular Genetics, 1111 Yata, Mishima, Shizuoka 411, Japan (E-mail:nfujita@dbj.nig.ac.jp, Tel:+81-559-75-0771(ex.684), Fax:+81-559-71-3651)
- REFERENCE 2 (sites)
AUTHORS An.G., Bendtak, D.S., Mamelak, L.A. and Friesen, J.D.
TITLE Organization and nucleotide sequence of a new ribosomal operon in Escherichia coli containing the genes for ribosomal protein S2 and elongation factor Ts
JOURNAL Nucleic Acids Res. 9 (16), 4163-4172 (1981)
MEDLINE 82059454
REFERENCE 3 (sites)
AUTHORS Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J.R.
TITLE The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the pyruvate dehydrogenase component
Eur. J. Biochem. 133 (1), 155-162 (1983)
JOURNAL 83209630
MEDLINE 4 (sites)
AUTHORS Stephens, P.E., Darlison, M.G., Lewis, H.M. and Guest, J.R.
TITLE The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the dihydroliipoamide acetyltransferase component
Eur. J. Biochem. 133 (3), 481-489 (1983)
JOURNAL 83234434
MEDLINE 5 (sites)
AUTHORS Stephens, P.E., Lewis, H.M., Darlison, M.G. and Guest, J.R.
TITLE Nucleotide sequence of the liipoamide dehydrogenase gene of Escherichia coli K12
Eur. J. Biochem. 135 (3), 519-527 (1983)
JOURNAL 84004369
MEDLINE 6 (sites)
AUTHORS Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J.C.
TITLE Regulation of expression and nucleotide sequence of the Escherichia coli gapD gene
J. Biol. Chem. 259 (23), 14824-14828 (1984)
JOURNAL 85054973
MEDLINE 7 (sites)
AUTHORS Broome-Smith, J.K., Edelman, A., Yousif, S. and Spratt, B.G.
TITLE The nucleotide sequences of the ponA and ponB genes encoding penicillin-binding protein 1A and 1B of Escherichia coli K12
Eur. J. Biochem. 147 (2), 437-446 (1985)
JOURNAL 85127060
MEDLINE 8 (sites)
AUTHORS Coulton, J.W., Mason, P., Cameron, D.R., Carmel, G., Jean, R. and Rode, H.N.
TITLE Protein fusions of beta-galactosidase to the ferrichrome-iron receptor of Escherichia coli K-12
- JOURNAL 165 (1), 181-192 (1986)
MEDLINE 86083668
REFERENCE 9 (sites)
AUTHORS Breton, R., Sanfacon, H., Papayannopoulos, I., Biemann, K. and Lapointe, J.
TITLE Glutamyl-tRNA synthetase of Escherichia coli. Isolation and primary structure of the gltX gene and homology with other aminoacyl-tRNA synthetases
J. Biol. Chem. 261 (23), 10610-10617 (1986)
JOURNAL 86278132
MEDLINE 10 (sites)
AUTHORS Koster, W. and Braun, V.
TITLE Iron hydroxamate transport of Escherichia coli: nucleotide sequence of the fhuB gene and identification of the protein
Mol. Gen. Genet. 204 (3), 435-442 (1986)
JOURNAL 87014116
MEDLINE 11 (sites)
AUTHORS Chye, M.L. and Pittard, J.
TITLE Transcription control of the aroP gene in Escherichia coli K-12: analysis of operator mutants
J. Bacteriol. 169 (1), 386-393 (1987)
JOURNAL 87083395
MEDLINE 12 (sites)
AUTHORS Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A. and Chang, S.
TITLE Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure
J. Bacteriol. 169 (2), 751-757 (1987)
JOURNAL 87109068
MEDLINE 13 (sites)
AUTHORS Coulton, J.W., Mason, P. and Allatt, D.D.
TITLE fhuC and fhuD genes for iron (III)-ferrichrome transport into Escherichia coli K-12
J. Bacteriol. 169 (8), 3844-3849 (1987)
JOURNAL 87279948
MEDLINE 14 (sites)
AUTHORS Tabor, C.W. and Tabor, H.
TITLE The speE operon of Escherichia coli. Formation and processing of a proenzyme form of S-adenosylmethionine decarboxylase
J. Biol. Chem. 262 (33), 16037-16040 (1987)
JOURNAL 88058963
MEDLINE 15 (sites)
AUTHORS Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wachter, E.
TITLE Complementary DNA and derived amino acid sequence of the precursor of one of the three protein components of the inter-alpha-trypsin inhibitor complex
FEBS Lett. 229 (1), 63-67 (1988)
JOURNAL 88152237
MEDLINE 16 (sites)
AUTHORS Andrews, S.C. and Guest, J.R.
TITLE Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12
Biochem. J. 255 (1), 35-43 (1988)
JOURNAL 89061679
MEDLINE 17 (sites)
AUTHORS Mellano, M.A. and Cooksey, D.A.
TITLE Nucleotide sequence and organization of copper resistance genes from Pseudomonas syringae pv. tomato
J. Bacteriol. 170 (6), 2879-2883 (1988)
JOURNAL 88227880
MEDLINE 18 (sites)
AUTHORS Liu, J.D. and Parkinson, J.S.
TITLE Genetics and sequence analysis of the pcnB locus, an Escherichia coli gene involved in plasmid copy number control
J. Bacteriol. 171 (3), 1254-1261 (1989)
JOURNAL 89155419
MEDLINE 19 (sites)
AUTHORS Sung, Y.C. and Fuchs, J.A.
TITLE Characterization of the cyn operon in Escherichia coli K12
J. Biol. Chem. 263 (29), 14769-14775 (1988)
JOURNAL 89008347
MEDLINE 20 (sites)

AUTHORS
TITLE
Lipinska, B., Sharma, S., and Georgopoulos, C.
Sequence analysis and regulation of the *htrA* gene of *Escherichia coli*: a sigma 32-independent mechanism of heat-inducible transcription
Nucleic Acids Res. 16 (21), 10053-10067 (1988)
89057448
21 (sites)
Roncero, M.I., Jepsen, L.P., Stroman, P., and van Heeswijk, R.
Characterization of a *leuA* gene and an *ARS* element from *Mucor circinelloides*
Gene 84 (2), 335-343 (1989)
90128278
22 (sites)
Xie, O.W., Tabor, C.W., and Tabor, H.
Spermidine biosynthesis in *Escherichia coli*: promoter and termination regions of the *speD* operon
J. Bacteriol. 171 (8), 4457-4465 (1989)
89327165
23 (sites)
Lindquist, S., Galleni, M., Lindberg, F., and Normark, S.
Signalling proteins in enterobacterial AmpC beta-lactamase regulation
Mol. Microbiol. 3 (8), 1091-1102 (1989)
90113890
24 (sites)
Kang, P.J., and Craig, E.A.
Identification and characterization of a new *Escherichia coli* gene that is a dosage-dependent suppressor of a *dnaK* deletion mutation
J. Bacteriol. 172 (4), 2055-2064 (1990)
90202727
25 (sites)
Nunn, D., Bergman, S., and Lory, S.
Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* *pili*
J. Bacteriol. 172 (6), 2911-2919 (1990)
90264276
26 (sites)
Martin-Verstraete, I., Debarbouille, M., Klier, A., and Rapoport, G.
Levanase operon of *Bacillus subtilis* includes a fructose-specific phosphotransferase system regulating the expression of the operon
J. Mol. Biol. 214 (3), 657-671 (1990)
90355183
27 (sites)
Surin, B.P., Watson, J.M., Hamilton, W.D., Economou, A., and Downie, J.A.
Molecular characterization of the nodulation gene, *nodI*, from two biovars of *Rhizobium leguminosarum*
Mol. Microbiol. 4 (2), 245-252 (1990)
90251164
28 (sites)
Honore, N., and Cole, S.T.
Query Match 12.8%; Score 178; DB 1; Length 8727;
Best Local Similarity 52.1%; Pred. No. 7.2e-27;
Matches 422; Conservative 0; Mismatches 385; Indels 3; Gaps 1;
Qy 382 ggctcatcatcagcaaaacggctacatcctgaccataaccacacgctgtgcggtatg 441
Db 69828 GTCATCATTGATGCGGATAAGGCTATGCTGCACCAACACACCGTTGTGTAACGCG 69887
Qy 442 ggcagatcaaaagtctgtctcaacgacgagcggaataataccgccaaactcctggtcg 501
Db 69888 ACGGTCATTAAGTTCAACTGAGCGATGCCCGTAAGTTCGACCGGAAGATGGTGCAAA 69947
Qy 502 gatgtccaatcagatgcgcctcttgaataatgacgacgacgaaagactaccgtctgc 561
Db 69948 GATCGCGGCTCTGATATCGCGTGTATCCAAATCCAGAACCCGAAAAACCTGACCGCAATT 70007
Qy 562 aaaaatcggaatcccaaaaatttgaacacgagcggaatgggtcgtccatcgcgcgcgc 621
Db 70008 AGATGGCGGATTCGATGCATTCGCGGTGGTGATTAACACCGATGGTAAACCGG 70067
Qy 622 ttgcgctttgacaacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 678

Db 70068 TTTGGTGTGGCGAGACGGTAACCTTCGGGATGTCTCTCGCGTGGGCGGTAGCGCCTG 70127
Qy 679 cccaaacgaagctacacaccccttcatccaaacccgacgctgcccataatccgggcaattcc 738
Db 70128 AATCGCGAAAACTACGAAAACTTTCATCCAGACCGATGACGGATCAACCGTGGTAACCTC 70187
Qy 739 ggcggccgcgtgtcaacttaaaagacagctgctgcgcatcaaatccgcaaatatacagc 798
Db 70188 GGTGGTGGCTGGTTAAACCTGAACGGCGAATGATCGGTATCAACACCGCGATCCTCGCA 70247
Qy 799 cgcgcgcgcggtattcatcgtggcatctcttgcaccccgattgacgttgccatgaatgct 858
Db 70248 CCGACGGCGGCAACATCGGTATCGGTTTGTCTATCCGAGTAACATGGTGAAAAACCTG 70307
Qy 859 gccaaacagctgaaaaacacccgcgcaaaagtccaaacgcgcgaactggggcgtgattatcag 918
Db 70308 ACCTCGCAGATGGTGAATACGGCAGGTGAACCGCGTGAGCTGGGTATTATGGGGACT 70367
Qy 919 gaagtatcctacggtttggcacagctgcttcggtctggataaaagccagcgccattgatt 978
Db 70368 GAGCTGAACCTCCGAACCTGGCGAAGCGATGAAGTTGACGCCACCGCGGTGCTTCGTA 70427
Qy 979 gccaaatccttcccggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1038
Db 70428 AGCAGGTTCTGCTAATTCCTCCGCTGCANAAAGCGGCGATTAAGCGGGTGATGTGATC 70487
Qy 1039 ctacgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1098
Db 70488 ACCTCACTGAACGGTAAGCCGATCAGCAGCTTTGCCGCACTGCGTCTCAGGTGGGTACT 70547
Qy 1099 attacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1158
Db 70548 ATGCGGTAGCAGCAAACTGACCCCTGGGCTTACTGCGGACGCTAAGCAGGTAAACGTG 70607
Qy 1159 aaagcgaagctggcgcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1188
Db 70608 AACCTGGNACTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 70637

RESULT 14
ECU70214
LOCUS 123171 bp DNA BCT 21-SEP-1996
DEFINITION
Escherichia coli chromosome minutes 4-6.
VERSION
U70214.1 GI:1552727
ACCESSION
U70214.1
KEYWORDS
Escherichia coli.
SOURCE
Escherichia coli.
ORGANISM
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 123171)
AUTHORS
Schramm, S., Duncan, M., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K., Federspiel, N., Hyman, R., Kalman, S., Komp, C., Kurdi, O., Lashkari, D., Lew, H., Lin, D., Namath, A., Oefner, P., Roberts, D., and Davis, R.W.
Sequence of minutes 4-25 of *Escherichia coli*
Unpublished
2 (bases 1 to 123171)
AUTHORS
Schramm, S., Duncan, M., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K., Federspiel, N., Hyman, R., Kalman, S., Komp, C., Kurdi, O., Lashkari, D., Lew, H., Lin, D., Namath, A., Oefner, P., Roberts, D., and Davis, R.W.
Direct Submission
Submitted (09-SEP-1996), Department of Biochemistry, Stanford University, Stanford, CA 94304, USA
FEATURES
Location/Qualifiers
1..123171
/organism="Escherichia coli"
/db_xref="taxon:562"
/map="4-6 min"
1..55116
/organism="Escherichia coli"
/note="AvrII fragment B"

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:51:56 ; Search time 120.64 Seconds
(without alignments)
2893.053 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 ggttcaaaaataccaata.....tcagggaagtcctccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	176.4	12.6	1854	1 V29528	Escherichia coli s
2	175.6	12.6	1980	1 Q1416	S. typhimurium htr
3	154.2	11.1	1791	1 T04403	Heat shock protein
4	139.4	10.0	2894	1 T10423	H. influenzae S833
5	128.8	9.2	1529	1 X14164	H. pylori GHFO 536
6	124	8.9	1540	1 V30845	Nucleotide sequenc
7	120.8	8.7	1484	1 V30921	Nucleotide sequenc
8	110.8	7.9	558	1 V31239	E. coli J96 pathog
9	105	7.5	2036	1 V29540	Homo sapiens D8725
10	105	7.5	2075	1 T75444	CDNA encoding oste
11	105	7.5	2205	1 V04680	Human presenilin 1
12	85	6.1	607	1 X29130	Polynucleotide RfP
13	71	5.1	284	1 V90593	Nucleotide sequenc
14	68	4.9	1553	1 V43618	Human secreted pro
15	62.8	4.5	273	1 V90690	Nucleotide sequenc
16	60.6	4.3	15747	1 X13018	Enterococcus faeca
17	60.6	4.3	21338	1 V52153	Streptococcus pneu
18	60.4	4.3	9848	1 X20503	Polynucleotide seq
19	57.6	4.1	1835	1 V29524	Homo sapiens PSP1
20	57.6	4.1	2040	1 V08860	Human cancer-relat
21	57.6	4.1	2187	1 V29535	Homo sapiens PSP1
22	57.6	4.1	2187	1 V29536	Homo sapiens PSP1
23	57.6	4.1	2187	1 V29539	Homo sapiens PSP1
24	57.4	4.1	1395	1 X34246	Mycobacterium spec
25	57.4	4.1	1771	1 T91473	Mycobacterium tube
26	57.4	4.1	1771	1 T91410	Mycobacterium tube
27	57.4	4.1	1771	1 V44351	Mycobacterium tube
28	57.4	4.1	1771	1 V64459	M. tuberculosis im
29	57.4	4.1	1812	1 X34247	Mycobacterium spec
30	57.4	4.1	2162	1 V37380	Streptococcus pneu
31	55	3.9	1506	1 X20639	Polynucleotide seq
32	51.6	3.7	1503	1 V29523	Homo sapiens PSP1
33	51.6	3.7	1787	1 V29522	Homo sapiens PSP1
34	51.6	3.7	1987	1 V74833	Staphylococcus aur

35 51.6 3.7 2144 1 V29538 Homo sapiens PSP1-
36 51.6 3.7 2551 1 V29537 Homo sapiens PSP1-
37 51.2 3.7 110000 1 X20248_07 Continuation (8 of
38 51.2 3.7 110000 1 X20248_08 Continuation (9 of
39 45.6 3.3 15518 1 X20521 Polynucleotide seq
40 41.4 3.0 1058 1 X34251 Mycobacterium spec
41 41.4 3.0 1068 1 X34251 Mycobacterium spec
42 41.4 3.0 1872 1 T91477 Mycobacterium tube
43 41.4 3.0 1872 1 T91414 Mycobacterium tube
44 41.4 3.0 1872 1 V44355 Mycobacterium tube
45 41.4 3.0 1872 1 V64463 M. tuberculosis im

ALIGNMENTS

RESULT 1

V29528
ID V29528 standard; cDNA; 1854 BP.
AC V29528;
DT 13-OCT-1998 (first entry)
DE Escherichia coli serine protease htrA.
KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease; htrA;
KW serine protease; neurodegeneration; predisposition; diagnosis; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
CDS 199..1674
FT /*tag= a
FT /product= htrA serine protease
FN EP-828003-A2.
PD 11-MAR-1998.
PR 26-AUG-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH,
PI Livi GP, Southern CD;
DI WPI; 98-161101/15.
DR P-PSDB; W56771.
PT Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
PS Example 2; Page 27-28; 65pp; English.
CC The sequence is that encoding the serine protease htrA which was used
CC in the isolation of PSP1.
SQ Sequence 1854 BP; 478 A; 470 C; 489 G; 417 T;

Query Match 12.6%; Score 176.4; DB 1; Length 1854;
Best Local Similarity 52.0%; Pred. No. 2e-38;
Matches 421; Conservative 0; Mismatches 386; Indels 3; Gaps 1;

QY 382 ggcttcacatcagcaaaacggcgatcctgacacataccacgctcgttgcggtatg 441
Db 547 GTCATCATTTGATGCGGATAAGGCTATGTCGTCACCAACACCGTGTGTGATAACGG 606
QY 442 ggcagatcaaaagtctgtctaacgacagcgaataataccgcaaacatcgcgttcg 501
Db 607 ACGGTCATTTAAAGTTCAACTGACGATGCCCGTAACTTCGACGCGAAGATGGTTGCAAA 666
QY 502 gatgtccaatccgagtgcctcttgaataatcgacgcaacggaagctaccctgcgtc 561
Db 667 GATCCCGGCTGTGATATCGCGTGTGATCCAAATCCAAACCCGAAACCTGACCGCAATT 726
QY 562 aaaaatcgcaatcccaaaaaattgaaacccggcggaatgggtcgctgcctatcgcgccgc 621
Db 727 AAGATGCGGATTTCTGATGACCTGCGCGTGGGTGATTACACCGTAGGATTTGTAACCCG 786
QY 622 ttccggtttgacaacacggctgaccgacgcatgttccgccaag---gcagaagcctg 678
Db 787 TTTGTGCTGGCGGAGACGGTAACCTCCGGGATTTGCTCTGCTGCTGGGGCGGTAGCGGCTG 846

CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma). 411 T;
SQ Sequence 1540 BP; 495 A; 280 C; 352 G; 411 T;

Query Match 8.9%; Score 124; DB 1; Length 1540;
Best Local Similarity 54.5%; Pred. No. 2.7e-24;
Matches 316; Conservative 0; Mismatches 255; Indels 9; Gaps 3;

QY 373 ttccgttcgggttcctcatcagcaaaacggtacatccctgacccaataccaccgctcgtt 432
DB 300 TTAGCAGCGGTGTCATCTTCTTAAGACGGCTATTTGTAACATAAACCATGTGATT 359
QY 433 gccggtatggcgagtcataaaagt---cctgctcaacgcaagcgcaatataccgcaaaa 489
DB 360 GATGCGCGGATTAAGATTAAAGTGACCATCCAGGAGCAATAAAGAAATATTCGCTACT 419
QY 490 ctcatcggttcggatgtccaatcccgatgtcgccctctgaaatcgacgcaacggaagag 549
DB 420 TTAGTAGCAGCGATCTGAAAGCGATTAGCCGTGATTTCGCATCA---CTAAGACAAAC 476
QY 550 ctaccgctgcgaataatcgcaatcccaaaatttgaaacccggcgcaatgggtcgctgcc 609
DB 477 TTGCCCACGATCAATCTCTGATTCTAAGCGATTTCAGTGGCGGATTTGGTTTTCG 536
QY 610 atcgcgccgcttcggtcttcaacacggtgacccggtcgatcgtgtccgc---caaa 666
DB 537 ATTGGTAACCCCTTTTGGCGTGGTGAAGCGTTACTCAAGGCAATTTTCAGCGCTCAAT 596
QY 667 ggcagaagcctgcggcgaagcgaacacacaccccttcacaaacgagtcggtgcaatcaat 726
DB 597 AAAAGCGGATTTGGGATCAACAGCTATGAGATTTTCAATCAACACAGCGCTCTAATAAT 656
QY 727 ccgggcaatcccgccggtcttcaactaaaggacagtcggtcggtacatcaatcg 786
DB 657 CTTGGAAATTTCCGGCGCGCTTTAATGTAGCCGTGGAGGGTTAGTGGGGATTAATACC 716
QY 787 caaatatagcgcgagcggcggtatccatgggcatctccttggcatcccgattgagctt 846
DB 717 GCTATCATCTCTAAACTTGGGGCAACACGCGATTTGGCTTGGCATCTCTTAACATG 776
QY 847 gccatgaatgcggcaagcgtgaaacacacgcaagtcgaatcccaacgcaactgggc 906
DB 777 GTTAAAGATATTGTAACCAACTCATCAAAACCGGTGAAGATTGAAGAGGTACTTGGCG 836
QY 907 gtgattattcagggaagtattcctacggtttggcacagtcgt 946
DB 837 GTGGGCTTGAAGATTGAGCGGCGATTGCAAAATTTCTT 876

RESULT 7
V90921
ID V90921 standard; DNA; 1484 BP.
AC V90921
DT 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone b8.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN W09849314-A2.
PN 05-NOV-1998.
PD 27-APR-1998; U08487.
PF 14-OCT-1997; US-061958.
PR 25-APR-1997; US-045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, McAtee CP;
WPI: 95-009433/01.
DR New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 37; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive

CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ Sequence 1484 BP; 472 A; 270 C; 344 G; 398 T;

Query Match 8.7%; Score 120.8; DB 1; Length 1484;
Best Local Similarity 54.1%; Pred. No. 1.9e-23;
Matches 314; Conservative 0; Mismatches 257; Indels 9; Gaps 3;

QY 373 ttccgttcgggttcctcatcagcaaaacggtacatccctgacccaataccaccgctcgtt 432
DB 299 TTAGCAGCGGTGTCATCTTCTTAAGACGGCTATTTGTAACATAAACCATGTGATT 358
QY 433 gccggtatggcgagtcataaaagt---cctgctcaacgcaagcgcaatataccgcaaaa 489
DB 359 GATGCGCGGATTAAGATTAAAGTGACCATCCAGGAGCAATAAAGAAATATTCGCTACT 418
QY 490 ctcatcggttcggatgtccaatcccgatgtcgccctctgaaatcgacgcaacggaagag 549
DB 419 TTAGTAGCAGCGATCTGAAAGCGATTAGCCGTGATTTCGCATCA---CTAAGACAAAC 475
QY 550 ctaccgctgcgaataatcgcaatcccaaaatttgaaacccggcgcaatgggtcgctgcc 609
DB 476 TTGCCCACGATCAATCTCTGATTCTAAGCGATTTCAGTGGCGGATTTGGTTTTCG 535
QY 610 atcgcgccgcttcggtcttcaacacggtgacccggtcgatcgtgtccgc---caaa 666
DB 536 ATTGGTAACCCCTTTTGGCGTGGTGAAGCGTTACTCAAGGCAATTTTCAGCGCTCAAT 595
QY 667 ggcagaagcctgcggcgaagcgaacacaccccttcacaaacgagtcggtgcaatcaat 726
DB 596 AAAAGCGGATTTGGGATCAACAGCTATGAGATTTTCAATCAACACAGCGCTCTAATAAT 655
QY 727 ccgggcaatcccgccggtcttcaactaaaggacagtcggtcggtacatcaatcg 786
DB 656 CTTGGAAATTTCCGGCGCGCTTTAATGTAGCCGTGGAGGGTTAGTGGGGATTAATACC 715
QY 787 caaatatagcgcgagcggcggtatccatgggcatctccttggcatcccgattgagctt 846
DB 716 GCTATCATCTCTAAACTTGGGGCAACACGCGATTTGGCTTGGCATCTCTTAACATG 775
QY 847 gccatgaatgcggcaagcgtgaaacacacgcaagtcgaatcccaacgcaactgggc 906
DB 776 GTTAAAGATATTGTAACCAACTCATCAAAACCGGTGAAGATTGAAGAGGTACTTGGCG 835
QY 907 gtgattattcagggaagtattcctacggtttggcacagtcgt 946
DB 836 GTGGGCTTGAAGATTGAGCGGCGATTGCAAAATTTCTT 875

RESULT 8
V31239/C
ID V31239 standard; DNA; 558 BP.
AC V31239
DT 01-OCT-1998 (first entry)
DE E. coli J96 pathogenicity island contig #53.
KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
KW PAI V; pheV; vaccine; protective immune response; ds.
OS Escherichia coli.
PN W09822575-A2.
PD 28-MAY-1998.
PF 21-NOV-1997; U21347.
PR 14-OCT-1997; US-061953.
PR 22-NOV-1996; US-031626.
PA (HUNA-) HUMAN GENOME SCI INC.
PA (UYWI-) UNIV WISCONSIN.

PI Choi GH, Dillon RJ, Welch RA;
DR WPI: 98-312461/27;
PT New isolated uropathogenic E. coli nucleotide sequences - used to
PT develop products for the detection of pathogenic E. coli and to
PT elicit an immune response to pathogenic E. coli
PS Claim 21: Page 144-145; 250pp; English.
CC This sequence represents a E. coli strain J96 contig containing
CC pathogenicity island (PAI) sequences, and represents a nucleic acid
CC molecule of the invention. PAIs are large fragments of DNA which comprise
CC pathogenicity determinants. The sequences of the invention are taken from
CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC approximately 94 min (at pheR) on the E. coli chromosome and is
CC approximately 160 kb in size. Antibodies specific to the proteins encoded
CC by the PAI open reading frames of the invention can be used in kits to
CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC a protective immune response in an animal to the uropathogenic E. coli
CC strain J96.
SQ Sequence 558 BP; 119 A; 156 C; 140 G; 138 T;

Query Match 7.9%; Score 110.8; DB 1; Length 558;
Best Local Similarity 55.6%; Pred. No. 6.8e-21;
Matches 269; Conservative 0; Mismatches 211; Indels 4; Gaps 3;

QY 382 ggttcacatcagcaaaacggctacatctcgtgacaaatcaccacgtgtgctggtatg 441
DB 513 GTCATCATGTATGCCGATAAGGCTATGTCGTCCACCAACACCCGCTGTGTATACNCG 454
QY 442 ggcagatcaaaagtcctcgaacgagcgcaatataccgccaaactcatcggttcg 501
DB 453 ACGGTCATTAAAGTTCACTGACGATGCCGTAAGTTCGACCGGAGATGGTTGGCAAA 394
QY 502 gatgtccaatccgatgctgccttctgaaatcgcagcaacggaagctaccctgctgc 561
DB 393 GATCCGCGCTCTGATATCGCGTGTATCCAAATCCAGAACCCGAAAAACCTGACCGCAATT 334
QY 562 aaaaatcggaatcccaaaaaatttgaaccggcggaatggctcgtccatcgcgccccc 621
DB 333 AGATGGCGGATCTGTATGCATCGCGTGGGTGATTACACCGTAGATGGTAACCCG 274
QY 622 ttccggtttgcaacagcgtgacgcgcgcgtcgtgctgcgcacaaaggcagaa--gcctgc 679
DB 273 TTTGGTCTGGCGAGACGGTAACCTTCGGGATTGTCCTCGCGTGGCGGTAGGGCTGA 214
QY 680 ccaacgaagctacacacccttcacaaacgcagctgttgccatcaaatccgggaattccg 739
DB 213 ATGCCGAAACTACGAAAACTTCATCCAGACCGATGCAGCGATCAACCGGTAACTCCG 154
QY 740 gcggcccgctgttcaacttaaaaggacagctgctggcatcaa--ttcgcaaatatacagc 798
DB 153 GTGGTGGCTGGTAAACCTGAACGGCGAACTGTATCGGTATCAACACCGCGATCCTCGCAC 94
QY 799 cgcagcggcggtatcattggcattctcttt--gcattcccgattgacgtttccatgaatgt 857
DB 93 CGGAGCGCGGCAACATCGGTATCNGTTTGGCTATCCCGAGTTAACTGTGTAANAACNT 34
QY 858 cggc 861
DB 33 GGAC 30

RESULT 9
V29540
ID V29540 standard; cDNA; 2036 BP.
AC V29540;
DT 13-OCT-1998 (first entry)
DE Homo sapiens D87258 sequence.
KW PS-1: presenilin; presenilin-1; PSP-1; Alzheimer's disease;
KW serine protease; neurodegeneration; predisposition; diagnosis;
KW D87258; ss.
OS Homo sapiens.
FH Key Location/Qualifiers

FT CDS 49..1491
FT /*tag= a
FT /product= D87258 protein
FT variation 1325
FT /*tag= b
FT /note= "G/T polymorphism"
PN EP-828003-A2.
PD 11-MAR-1998.
PF 26-AUG-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIK) SMITHKLINE BECHAM CORP.
PA (SMIK) SMITHKLINE BECHAM PLC.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH,
PI Livi GP, Southern CD;
DR WPI: 98-161101/15.
DR P-PSDB; W56778.
PT Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
PS Claim 5; Page 30; 65pp; English.
CC The sequence is that of an isolated cDNA clone D87528 which can be used
CC to identify modulators of serine protease activity and also to diagnose
CC a condition associated with lack of one of the serine proteases
CC or a genetic predisposition to neurodegeneration in a patient,
CC preferably predisposition to Alzheimer's disease.
SQ Sequence 2036 BP; 451 A; 581 C; 590 G; 413 T;
Query Match 7.5%; Score 105; DB 1; Length 2036;
Best Local Similarity 55.4%; Pred. No. 4e-19;
Matches 232; Conservative 0; Mismatches 175; Indels 12; Gaps 1;
QY 376 ggttcgggcttcacatcagcaaaacggctacatcctcgacaaatcaccacgtgtgccc 435
DB 658 GGGTCTGGGTTATTGTCTCGAAGATGGACTGATCGTGACAAATGCCACGCTGGTGACC 717
QY 436 ggtatggcgagtcacaaagtcctcacaacgacgaagcgcaatataccgcgaacatcacc 495
DB 718 AACAAAGCACCAGGGTCAAAGTTGAGCTGAAGAACGGTGCACCTTACGAAGCCAAATCAAG 777
QY 496 ggttcgagtcacaaatcgatgctgccttcgaaatcgcagcaacgagagatcacc 555
DB 778 GATGTGATGAGAAGCAGACATCGACTCATCAAAATTGACCCACGAGGACGCTGCCT 837
QY 556 gtctgcaaaatcggaatcccaaaatttgaaacggcgcaatggctgctgcatcgcc 615
DB 838 GTCTGTGCTGTGGCCGCTCTCTCAGAGCTGGCGCGGAGAGTTCTGTGCTGCCATCGGA 897
QY 616 gcgccttcggcttgacacacagcgtgacgcgcgcgcgtgctccgcgaaggcggaagc 675
DB 898 AGCCCGTTTTCCCTTCAAAACACAGTCACCCACCGGATCGTGAGCACCACCCAGCGAGC 957
QY 676 -----ctgcccacgaaagctacacaccccttcacaaacgaggttgcctc 723
DB 958 GGCAAAAGCTGGGGCTCCCGCAACTCAGACATGAGCTACATCCAGACCGACCGCATCATC 1017
QY 724 aatccgggcaatccggcgcccgctgttcaacttaaaagacaggtgctcgggcatcaa 782
DB 1018 AACTATGAAACTCGGGAGGCCCGTTAGTAAACCTGGACGGTGAAGTGATTGGAATTAA 1076
RESULT 10
T75444
ID T75444 standard; cDNA; 2075 BP.
AC T75444;
DT 16-SEP-1997 (first entry)
DE cDNA encoding osteoblast like cell derived protein.
KW Osteoblast like cell; prevention; treatment; disease;
KW contraceptive; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 88..1530

```
FT signal_peptide /*tag= a
FT 88..168
FT /*tag= b
FT mat_peptide 169..1527
FT /*tag= c
PN J09107980-A.
PD 28-APR-1997.
PF 12-AUG-1996; 231415.
PR 17-AUG-1995; JP-233537.
PA (NLSB ) JAPAN TOBACCO INC.
DR WPI; 97-292469/27.
DR P-PSDB; W22849.
PT DNA encoding osteoblast like cell derived protein - useful for
PT treatment and prevention of various diseases and as contraceptive
PS Claim 1; Pages 21-23; 42pp; Japanese.
CC The present sequence encodes an osteoblast like cell derived
CC protein, which may be used to prevent or treat various diseases, or
CC as a contraceptive.
CC mRNA was collected from an osteoblast like cell, and double
CC stranded DNA synthesised. A 3'-directed cDNA library was prepared,
CC and the base sequence of the DNA of each clone in the cDNA library
CC determined. An osteoblast complete chain cDNA library was prepared,
CC and the clone GS2422, which has a complete chain, isolated. GS2422
CC protein was expressed and purified, and an antibody against GS2422
CC prepared. The amino acid sequence of a purified specimen of GS2422
CC protein was analysed. The GS2422 gene is expressed in various human
CC organs.
SQ Sequence 2075 BP; 459 A; 592 C; 603 G; 421 T;

Query Match 7.5%; Score 105; DB 1; Length 2075;
Best Local Similarity 55.4%; Pred. No. 4.1e-19;
Matches 232; Conservative 0; Mismatches 175; Indels 12; Gaps 1;

QY 376 ggttcgggttcacatcagcaaaacggctacatccttgacaaataccacacgtctgtgcc 435
DB 376 ggttcgggttcacatcagcaaaacggctacatccttgacaaataccacacgtctgtgcc 435
QY 436 ggtatgggcagtatcaaaagtcctgtcacaacgacgagcggaatataccgccaaactcacc 495
DB 436 ggtatgggcagtatcaaaagtcctgtcacaacgacgagcggaatataccgccaaactcacc 495
QY 496 ggttcggatgtccaatccgatgtgccctcttgaaatcgacgcaacgagagctacc 555
DB 496 ggttcggatgtccaatccgatgtgccctcttgaaatcgacgcaacgagagctacc 555
QY 817 GATGTGGATGAGAAAGACAGATCGCATCAAAATTTGACCAACCGAGCGTGCCT 876
DB 817 GATGTGGATGAGAAAGACAGATCGCATCAAAATTTGACCAACCGAGCGTGCCT 876
QY 556 gtccgcaaaatcggcaatcccaaaatttgaaacggcggaatgggtcgtgcctcgc 615
DB 556 gtccgcaaaatcggcaatcccaaaatttgaaacggcggaatgggtcgtgcctcgc 615
QY 877 GTCTGCTGCTTGGCGCTCCTCAGAGCTGCGGCGGGAGAGTTCGTGCGCATCGGA 936
DB 877 GTCTGCTGCTTGGCGCTCCTCAGAGCTGCGGCGGGAGAGTTCGTGCGCATCGGA 936
QY 616 ggcgccttcgggtttgacaaacgctgacccggtcgtcgtccgcgaagcagaagc 675
DB 616 ggcgccttcgggtttgacaaacgctgacccggtcgtcgtccgcgaagcagaagc 675
QY 937 AGCCCGCTTTTCCTTCAAAACACAGTCAGCACCGGGATCGTGAGCACCCACCGAGCGGC 996
DB 937 AGCCCGCTTTTCCTTCAAAACACAGTCAGCACCGGGATCGTGAGCACCCACCGAGCGGC 996
QY 676 -----ctgcccacgaagctacacaccccttcacaaacacacgttgccatc 723
DB 676 -----ctgcccacgaagctacacaccccttcacaaacacacgttgccatc 723
QY 997 GCAAAAGAGCTGGGGTTCGCAACTCAGACATGGACTACATCCAGCCAGCGCCATCATC 1056
DB 997 GCAAAAGAGCTGGGGTTCGCAACTCAGACATGGACTACATCCAGCCAGCGCCATCATC 1056
QY 724 aatccgggcaattccggcgcccgctgttcaacttaaaagcagagtcgtcggcatcaa 782
DB 724 aatccgggcaattccggcgcccgctgttcaacttaaaagcagagtcgtcggcatcaa 782
QY 1057 AACTATGGAAACTCGGGAGCCCGCTTAGTAAACCTGGACGGTGAAGTATTGGAATTAA 1115
DB 1057 AACTATGGAAACTCGGGAGCCCGCTTAGTAAACCTGGACGGTGAAGTATTGGAATTAA 1115

RESULT 11
V04680
ID X29130 standard; cDNA; 2205 BP.
AC V04680;
DT 20-JUL-1998 (first entry)
DE Human presenilin interacting protein gene mutTMI-TM2.
KW Presenilin interacting protein; mutTMI-TM2; serine protease;
KW heat shock protein; human; familial Alzheimer's disease; FAD;
KW diagnosis; therapy; ss.
KW Homo sapiens.
```

```
PN WO9801549-A2.
PD 15-JAN-1998.
PF 04-JUL-1997; CA0475.
PR 02-JAN-1997; US-034590.
PR 05-JUL-1996; US-021673.
PR 12-JUL-1996; US-021700.
PR 08-NOV-1996; US-029895.
PA (HSCR-) HSC RES & DEV LP.
PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.
PI Fraser PE, Rommens JM, St George-Hyslop PH;
DR WPI; 98-286355/25.
PT New isolated mutant presenilin-1 genes - useful for developing
PT products for use in detection, diagnosis and therapy of Alzheimer's
PT disease and for drug screening.
PS Claim 4; Page 222-223; 238pp; English.
CC This DNA sequence, mutTMI-TM2, codes for an unknown protein that
CC interacts with presenilin proteins (see W23964-67). To identify
CC such proteins (see also V04674-80), a yeast two-hybrid system was
CC used to screen a brain cDNA library for clones which interacted
CC with functional domains of the presenilins. The mutTMI-TM2 clone
CC interacts with a mutant presenilin-1 transmembrane 1-2 loop domain,
CC and appears to correspond to a known heat-shock serine protease
CC gene. Methods for identifying substances that affect the
CC interaction of a presenilin-interacting protein with presenilin are
CC provided. Such substances may be useful in the treatment of
CC Alzheimer's disease. Transgenic animals carrying the mutTMI-TM2
CC nucleic acid, pure preparations of the mutTMI-TM2 protein, and
CC antibodies that selectively bind mutTMI-TM2 protein are also
CC claimed.
SQ Sequence 2205 BP; 525 A; 609 C; 614 G; 439 T;

Query Match 7.5%; Score 105; DB 1; Length 2205;
Best Local Similarity 55.4%; Pred. No. 4.2e-19;
Matches 232; Conservative 0; Mismatches 175; Indels 12; Gaps 1;

QY 376 ggttcgggttcacatcagcaaaacggctacatccttgacaaataccacacgtctgtgcc 435
DB 376 ggttcgggttcacatcagcaaaacggctacatccttgacaaataccacacgtctgtgcc 435
QY 436 ggtatgggcagtatcaaaagtcctgtcacaacgacgagcggaatataccgccaaactcacc 495
DB 436 ggtatgggcagtatcaaaagtcctgtcacaacgacgagcggaatataccgccaaactcacc 495
QY 496 ggttcggatgtccaatccgatgtgccctcttgaaatcgacgcaacgagagctacc 555
DB 496 ggttcggatgtccaatccgatgtgccctcttgaaatcgacgcaacgagagctacc 555
QY 854 GATGTGGATGAGAAAGACAGATCGCATCAAAATTTGACCAACCGAGCGTGCCT 913
DB 854 GATGTGGATGAGAAAGACAGATCGCATCAAAATTTGACCAACCGAGCGTGCCT 913
QY 556 gtccgcaaaatcggcaatcccaaaatttgaaacggcggaatgggtcgtgcctcgc 615
DB 556 gtccgcaaaatcggcaatcccaaaatttgaaacggcggaatgggtcgtgcctcgc 615
QY 914 GTCTGCTGCTTGGCGCTCCTCAGAGCTGCGGCGGGAGAGTTCGTGCGCATCGGA 973
DB 914 GTCTGCTGCTTGGCGCTCCTCAGAGCTGCGGCGGGAGAGTTCGTGCGCATCGGA 973
QY 616 ggcgccttcgggtttgacaaacgctgacccggtcgtcgtccgcgaagcagaagc 675
DB 616 ggcgccttcgggtttgacaaacgctgacccggtcgtcgtccgcgaagcagaagc 675
QY 974 AGCCCGCTTTTCCTTCAAAACACAGTCAGCACCGGGATCGTGAGCACCCACCGAGCGGC 1033
DB 974 AGCCCGCTTTTCCTTCAAAACACAGTCAGCACCGGGATCGTGAGCACCCACCGAGCGGC 1033
QY 676 -----ctgcccacgaagctacacaccccttcacaaacacacgttgccatc 723
DB 676 -----ctgcccacgaagctacacaccccttcacaaacacacgttgccatc 723
QY 1034 GCAAAAGAGCTGGGGTTCGCAACTCAGACATGGACTACATCCAGCCAGCGCCATCATC 1093
DB 1034 GCAAAAGAGCTGGGGTTCGCAACTCAGACATGGACTACATCCAGCCAGCGCCATCATC 1093
QY 724 aatccgggcaattccggcgcccgctgttcaacttaaaagcagagtcgtcggcatcaa 782
DB 724 aatccgggcaattccggcgcccgctgttcaacttaaaagcagagtcgtcggcatcaa 782
QY 1094 AACTATGGAAACTCGGGAGCCCGCTTAGTAAACCTGGACGGTGAAGTATTGGAATTAA 1152
DB 1094 AACTATGGAAACTCGGGAGCCCGCTTAGTAAACCTGGACGGTGAAGTATTGGAATTAA 1152

RESULT 12
X29130
ID X29130 standard; cDNA; 607 BP.
AC X29130;
DT 04-JUN-1999 (first entry)
DE Polynucleotide RfP633 having hypoxia regulated activity.
KW Hypoxia; angiogenesis; apoptosis regulated activity; ischemia; tumour;
KW hypoxic retina; ss.
KW Homo sapiens.
```

```
OS Rattus sp.
PN WO909046-A1.
PD 25-FEB-1999.
PF 21-AUG-1997; U17297.
PR 21-AUG-1997; US-056453.
PA (KOHN/) KOHN K I.
PA (QUAR-) QUARK BIOTECH INC.
PI Einat P. Skalliter R.
DR WPI; 99-180964/15.
PT New polynucleotides with hypoxia, angiogenesis and apoptosis
PT regulated activity - useful for diagnosis of ischemia, and for
PT treatment involving gene therapy
PS Claim 1: Page 56-57; 72pp; English.
CC Sequences 229126 to 229136 represent isolated polynucleotides with
CC hypoxia, angiogenesis and/or apoptosis regulated activity. The invention
CC provides a method for diagnosing ischemia from bodily fluid samples by
CC identifying at least one expressed gene, or identifying at least one up-
CC regulated gene in a tissue sample. In situ analysis was performed for the
CC above sequences in solid tumours and hypoxic retinas. The expression of
CC the genes is activated in tumours around the hypoxic region in the tumour
CC centre, therefore are hypoxia-regulated in vivo. Up regulation promotes
CC angiogenesis required for tumour growth. The genes, proteins,
CC antigens, antisense oligonucleotides and dominant negative peptides
CC are useful for regulating angiogenesis, apoptosis and hypoxia.
SQ Sequence 607 BP; 157 A; 148 C; 166 G; 134 T;

Query Match 6.1%; Score 85; DB 1; Length 607;
Best Local Similarity 53.9%; Pred. No. 6.6e-14;
Matches 226; Conservative 0; Mismatches 180; Indels 13; Gaps 2;

QY 376 ggtcggcttcacatcagcaaaacggctacatctcgacaaatccacgcgtgtgccc 435
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 GGTCAGGATTCATTGTGCGGAGGATGGACTGATTGTGCAAAATGCTCAGTGGTGACC 164
QY 436 ggtatggcagtagtaagtcctgctcaacgacagcggaatacgcgcaaacctcatc 495
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 AACAAAAACCGGTCAGGTTGGCTCAAGAAATGAGCGGACTTATGAAGCCAAATCAAG 224
QY 496 ggtcggatgcacatcagatgtgcccctctgaaatcgacgcaagagctacccc 555
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 GATGTGGATGAAAAGGCGCATTTGCGCTTATCAAGATTGACCAACGAGGGTAAAGTGCCA 284
QY 556 gtctgcaaaatcgcaatcccaaaattgaaacggcgcaaatgggtcgctgcccacgc 615
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 GTCTGTGCTTGGCGCTCTCGAGCTGGCGCCAGGAGAAATTTGTGTTGCCATCGGA 344
QY 616 gcgccccttggtttgacaacagcgtgaccgcccgcacgtgctccgcaaaaggcagaagc 675
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 AC-CCCTTCTCTTCAAAACACGGTCACCACTGGGATCGTCAGCACCAACCCACGGAGGC 403
QY 676 -----ctgcccacgaaagctacacaccccttcacccaaacgacggttgcacac 723
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 GGCAGAAAGAACTGGGGCTCCGGAACCTCGGATATGGACTACATTTCAGACAGAGCCCATC 463
QY 724 aatccgggcaattccggcgcccgctgttcaacttaaaagacaggtcgctcgccatcaa 782
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 AATTATGGAACCTCCGAGGAGCCGCTTAGTAACCTGGATGGCGAGGTGATTGGGATAA 522

RESULT 13
V90593/C
ID V90593 standard; DNA; 284 BP.
AC V90593;
DE 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone Y107-173.SEQ from cluster 4.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN WO9849314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
```

```
PR 25-APR-1997; US-045107.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, McAttee CP;
DR WPI; 99-009433/01.
PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 124; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ Sequence 284 BP; 75 A; 69 C; 60 G; 80 T;

Query Match 5.1%; Score 71; DB 1; Length 284;
Best Local Similarity 56.1%; Pred. No. 3.1e-10;
Matches 134; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 682 aacgaagctacacacccttcacaaacggcagcttgccatcaatccgggcaattccggc 741
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 ATCAACAGCATATGAGAAATTTTCATCAACAGACAGCGCTCTATTAAATCTGGAATTCGGC 184
QY 742 gcccgcgtgttcaacttaaaaggacagctgcgcgcatcaatccgaaatatacacgcc 801
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 GCGCTTAAATGTAGTCGCGGTAGTGGGGATTAAATACCGCTATCATCTCTATAA 124
QY 802 agcggcggtatcatgggcatctcctttgccatcccgattgacgttgccatgaatgcgcc 861
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 ACTGGGGCAACCCAGCGCATTTGGCTTTGCCATCCCTTCAACATGGTTAAAGATATTGTA 64
QY 862 gaacagctgaaaaaacacccggcaaaagtcacacgcggaactggggcgtgatttcagg 920
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
63 ACCCAACTACTCAAAACCGGTAAAGATTGAAGAGGTTACCTGGCGCTGGCTTGCAGA 5

RESULT 14
V43618
ID V43618 standard; DNA; 1553 BP.
AC V43618;
DE 24-SEP-1998 (first entry)
DE Human secreted protein 18 encoding DNA.
KW Secreted protein; human; cell proliferation; cytokine activity;
KW tissue growth; cellular differentiation; regeneration; activin;
KW inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW anti-inflammatory activity; biomarker; ss.
OS Homo sapiens.
PI Key Location/Qualifiers
PI CDS 65..1495
PI /*tag= a
PI /product= "human secreted protein"

WO9825959-A2.
PD 18-JUN-1998.
PF 11-DEC-1997; U22787.
PR 11-DEC-1996; US-032757.
PA (CHIR ) CHIRON CORP.
PI Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR WPI; 98-348453/30.
DR P-PSDB: W63698.
DR Secreted human polypeptides - having cytokine, cell proliferation or
DR differentiation, activin or inhibin, tumour inhibition or
PT anti-inflammatory activities
PS Claim 6; Pages 45-46; 78pp; English.
CC This DNA encodes a human secreted protein. The specification provides
CC secreted protein sequences (W63681 to W63699) encoded by the nucleic
CC acid sequences shown in V43601 to V43619. The invention provides a
```

CC method of identifying a secreted polypeptide which is modified by rough
CC microsomes. The secreted proteins can be used in assays to determine
CC biological activities, such as cytokine, cell proliferation, or cellular
CC differentiation activities, tissue growth or regeneration, activin or
CC inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC anti-inflammatory activity. The proteins can also be used as biomarkers,
CC to identify tissues or cell types which express the proteins, or a stage-
CC or disease-specific alteration in protein expression. They can be used
CC in protein interaction assays, to identify ligands or binding proteins.
CC Compounds which affect the biological activities of the secreted proteins
CC or their ability to interact with specific ligands can be identified
CC using the proteins in screening assays. The proteins and antibodies that
CC bind specifically to the protein can also be used to design diagnostic
CC tests and therapeutic compositions for diseases which may be associated
CC with altered expression of these proteins. Fusion proteins comprising,
CC e.g. signal sequences or transmembrane domains of the proteins can be
CC used to target other protein domains to cellular membrane or they can
CC be secreted extracellularly.
SQ Sequence 1553 BP; 348 A; 384 C; 467 G; 354 T;

Query Match 4.9%; Score 58; DB 1; Length 1553;
Best Local Similarity 49.5%; Pred. No. 3.7e-09;
Matches 213; Conservative 0; Mismatches 205; Indels 12; Gaps 1;
Qy 368 tgaacttcggttcggttcacatcagcaaaacggtacatcctcgccaataccacg 427
Db 660 TGTACAGTGGCTCTGGGTTCAATAGTCTGTGAGGAGGGCTCATTTATCCCAATGCCCATG 719
Qy 428 tcgttcggttatggcagatcaaaagtcctgctcaacgacaaagcggaatataccgcca 487
Db 720 TTGTCAGGAACAGCAGTGGATTGAGTGGTGTCTCCAGAATGGGGCCGCTTATGAAGCTG 779
Qy 488 aactcatcggttcggtatccaatcgatgtcgccctctgaaatcgacgcaacggaag 547
Db 780 TTGTCAGGATATTGACCTAAATTTGGATCTTGGCGTGAATGAAGATTGAATCAATGCTG 839
Qy 548 agctaccgctgcataaatcggcaatcccaaaattgaaacggcggaatgggtcgctg 607
Db 840 AACTTCCTGATCTGATGCTGGGAGATCATCTGACCTCGGGCTGGAGAGTTGTGGTGG 899
Qy 608 ccacggtcgcccttcggttcgttgaaacagcgtgacccggtcggtcggtcgccaaag 567
Db 900 CTTTGGCAGGCCATTTTCTCTGCAGACACACAGCTACTCAGGAATGTCAGCACCAAC 959
Qy 668 gcagaag-----ccctgcccaacgaaagctacacaccccttcacaaacgcgcg 715
Db 960 AGCGAGGGGGCAAGAACTGGGGATGAAGGATTCAGATATGGACTACGTCCAGATTGATG 1019
Qy 716 ttgccatcaatccgggaattccggttcggttcgttcgaactaaagacaggtcgctg 775
Db 1020 CCACATTAATATGGAATTTCTGTGGTGGTCTCTGTGGAATCTGGATGGTGTGATG 1079
Qy 776 gcatcaattc 785
Db 1080 GGTCAATTC 1089

RESULT 15
V90690
ID V90690 standard; DNA; 273 BP.
AC V90690;
DT 18-FEB-1999 (first entry)
DE Nucleotide sequence of clone Y261T3.SEQ from cluster 4.
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
PN W0949314-A2.
PD 05-NOV-1998.
PF 27-APR-1998; U08487.
PR 14-OCT-1997; US-061958.
PR 25-APR-1997; US-045107.

PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Chow TP, Fry KE, Lim MY, Mcatee CP;
DR WPI; 99-009433/01.
PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
PS Claim 27; Page 155; 402pp; English.
CC The present sequence encodes a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ Sequence 273 BP; 78 A; 58 C; 64 G; 73 T;

Query Match 4.5%; Score 62.8; DB 1; Length 273;
Best Local Similarity 54.9%; Pred. No. 5e-08;
Matches 124; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Qy 682 aacgaagctacacacccttcacacacacgacgttgccatcaatccggcaattccgac 741
Db 48 ATCAACAGCTATGAGATTTTCATTCAAAACAGACGCTCTATTAATCTCTGGAATTCGCGC 107
Qy 742 gcccgcgtgtcaacttaaaagacagctcgtcggcatcaattcgcaaatatacacgccc 801
Db 108 GCGCTTTAATGTAGCCGTGGAGGTTAGTGGGATTAATACCGCTATCATCTCTAAA 167
Qy 802 agcggcggtatccatcggttccttggccatcccgatgacggtgcccgaatgctgcc 861
Db 168 ACTGGGGGCAACCCACGCGCATTTGGCTTTGCCATCCCTTCTAACATGTTAAAGATATTGA 227
Qy 862 gaacagctgaaaaaacacacggcaaaagtcacacgacgacgaactggcg 907
Db 228 ACCCAACTCATCAAAACCGGTTAGATTGAAGAATTACTTGGGCG 273

Search completed: April 14, 2000, 16:52:18
Job time: 6877 sec

Tue Apr 18 14:25:00 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2000, 16:28:42 ; Search time 83.49 Seconds
(without alignments)
2000.405 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgtcaaaaataccaata.....tcagggaagtcgccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCUT9_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175.6	12.6	1980	2	US-08-350-741-1
2	175.6	12.6	1980	4	US-08-463-875A-1
3	154.2	11.1	1791	1	US-08-245-294-7
4	154.2	11.1	1791	2	US-08-474-499-7
5	154.2	11.1	1791	2	US-08-307-279A-7
6	154.2	11.1	1791	6	PCT-US95-06211-7
7	139.4	10.0	2894	1	US-08-278-091-1
8	139.4	10.0	2894	1	US-08-483-859-1
9	139.4	10.0	2894	1	US-08-472-173-1
10	139.4	10.0	2894	3	US-08-487-167-1
11	139.4	10.0	2894	4	US-08-482-816-1
12	139.4	10.0	2894	4	US-08-296-149-1
13	139.4	10.0	2894	4	US-08-801-499-1
14	139.4	10.0	2894	4	US-08-615-271-1
15	136.2	9.8	1611	1	US-08-485-569-1
16	136.2	9.8	1611	1	US-08-480-993-1
17	136.2	9.8	1611	3	US-07-903-079B-1
18	85.8	6.2	539	1	US-08-322-742-18
19	41	2.9	1187	2	US-08-440-856A-2
20	39.2	2.8	2414	7	5248599-1
21	37.2	2.7	2793	2	US-08-209-747-1
22	37.2	2.7	2793	2	US-08-458-298-1
23	36.8	2.6	716	4	US-08-211-718-1
24	36.8	2.6	1608	4	US-08-211-718-8
25	35.8	2.6	2167	3	US-08-461-775-9
26	35.8	2.6	2668	3	US-08-461-775-11
27	35.6	2.6	424	4	US-08-476-176B-7

Sequence 2, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 10, Appli
Sequence 10, Appli
Sequence 185, App
Sequence 185, App
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 1, Appli

28 35.6 2.6 1560 4 US-08-709-979A-2
29 35.4 2.5 783 1 US-08-264-861A-11
30 35.4 2.5 783 6 PCT-US95-07784-11
31 35.4 2.5 1150 1 US-08-264-861A-10
32 35.4 2.5 1150 6 PCT-US95-07784-10
33 35.2 2.5 898 4 US-08-997-080-185
34 35.2 2.5 898 4 US-08-997-362-185
35 34.8 2.5 28958 1 US-08-258-261B-6
36 34.8 2.5 28958 1 US-08-456-837-6
37 34.8 2.5 28958 1 US-08-457-342-6
38 34.8 2.5 28958 1 US-08-457-646A-6
39 34.8 2.5 28958 2 US-08-458-076A-6
40 34.8 2.5 28958 2 US-08-764-233A-4
41 34.8 2.5 28958 2 US-08-457-335A-6
42 34.8 2.5 28958 2 US-08-729-214-6
43 34.8 2.5 49377 2 US-08-764-233A-1
44 34.6 2.5 833 3 US-08-403-852D-3
45 34.6 2.5 5392 3 US-08-403-852D-1

ALIGNMENTS

RESULT 1
US-08-350-741-1
; Sequence 1, Application US/08350741
; Patent No. 5804194
; GENERAL INFORMATION:
; APPLICANT: DOUGAN G.,
; APPLICANT: CHARLES I.G.,
; APPLICANT: HORMACHE C.E.,
; APPLICANT: JOHNSON K.S.,
; APPLICANT: CHATFIELD S.N.
; TITLE OF INVENTION: LIVE VACCINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON and VANDERHYE PC
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/350,741
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 09-MAY-1994
; APPLICATION NUMBER: US 07/952,737
; FILING DATE: 30-NOV-1992
; APPLICATION NUMBER: GB 9007194.5
; FILING DATE: 30-MAR-1990
; APPLICATION NUMBER: PCT/GB91/00484
; FILING DATE: 28-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double


```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1652
US-08-307-279A-7
```

```

Query Match      11.1%; Score 154.2; DB 2; Length 1791;
Best Local Similarity 50.4%; Pred. No. 1.5e-33;
Matches 403; Conservative 0; Mismatches 393; Indels 3; Gaps 1;

QY 373 ttcgggttcgggttcacatcagcaaaaacggctcatcctcctgaccaaataccacgcgttc 432
DB 516 TTGGATCGGGTTTTTTTATCTCGTCTGATGTTATTTGTGACCAATAATCATGTGATT 575

QY 433 gccgggtatgggcagatcaaaagtcctcgtcaacgcaagcgcgcaataaccgcaaaactc 492
DB 576 TCTGATGGCACAAGTTACGCTGTGTTCTTGATGACGGTACAGAACTGAATGCAAAACTC 635

QY 493 atcgggttcggatgccaatccgatgcgccttcctgaaaaatcgacgcaacggaagagcta 552
DB 636 ATTGAACGGACCCACGACACTGATCTTCAGAGTATTAAAGTCAATGCAAAAAGAAAATTT 695

QY 553 cccgttcgtaaaatcggcaatcccaaaaatttgaacccggcggaatgggtcgctgccatc 612
DB 696 TCGTACGTTGATTTGGTGATGATTCAAAACTTCGTTGGTGATTTGGGTTTGTGCTATT 755

QY 613 ggcgcgccttcgggttcgacaaagtcgacgctgcgccttcgaaaaatcgacgcaacggaagcaga 672
DB 756 GGTAATCCATTTGGTCTTGGTGAACCTGACAGAGGTATCGTTTCAGCACGTGGACGT 815

QY 673 agcctgcccacgaagctcacacaccttcacccaacccgagcttgcccatcaatccgggc 732
DB 816 GATATCGGTACCGGTTTATGATGATTTATTCAGATTGATGCTGCAGTTAATCGAGGA 875

QY 733 aattccggcgccgcgtgttcaacttaaaagacaggttcgctgcgccttcgaatcgcaaaata 792
DB 876 AATTCTGGAGGTCCAACCTTTTGTATCTTAACGGAAAGGTTTGGAGTGAATACGGCAATT 935

QY 793 tacagccgacgcgcggtattcggcatctcctcttgccatcccgattgacgttgcctatg 852
DB 936 TTTTCTCTCTTCGTTGGGCAACGTTGGGATGCTTTTCGCTATTCGGCAGCAACAGCGAAC 995

QY 853 aatgtcgcgaacagctgaaaaacacccgcaaaagtcctcctcctcctcctcctcctcctcctc 912
DB 996 GAGGTTGTGCAACAACCTTATCGAAAAAGGTTTAGTTTCAGCGTGGTTGGGTTTCAG 1055

QY 913 attcagaagatcctacgtgttggcacagctcgttcggctcgatataaagccagcgcgca 972
DB 1056 ATTACGCTGTAAACAAAAGAAATTTCTGATTTCAATTTGGTTTGAAGGAGGCTAAGGTGCG 1115

QY 973 ttgattgcaaaaatcctcccgagcccgagcccgagaaagtcgctgcggcctgcaggcgcgac 1032
DB 1116 TTAATTAACCATCCATTAAAGGGCCAGCGCGCAAAA--GCTGGTATCAAGCGAGTGAT 1172

QY 1033 atcgtcctcagctcagcgcgcgagaaatacgttcttcctccggcgaccttcctcgtaagtc 1092
DB 1173 GTTATTATTTTCGGTAAATGTTGAGAGATTAAATGATGTCCTGATCTAGCAAAAGCGTATT 1232

QY 1093 ggcgcattacgccccgaaagaaagtcagctcggcgtatggcgcaaaagcgaaagaaatc 1152
DB 1233 GCAAAATATGAGCCAGGAGAAACAGTAACCTTAGGAGTTTGGAAATCTGTTAAAGAGAG 1292

QY 1153 acaatcaaaagccaagctgg 1171
DB 1293 AATATTAAGTTAACTTG 1311
```

RESULT 6

PCT-US95-06211-7

; Sequence 7, Application PC/TUS9506211

; GENERAL INFORMATION:

```

; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06211
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18 MAY 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.6121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1649
PCT-US95-06211-7
```

```

Query Match      11.1%; Score 154.2; DB 6; Length 1791;
Best Local Similarity 50.4%; Pred. No. 1.5e-33;
Matches 403; Conservative 0; Mismatches 393; Indels 3; Gaps 1;

QY 373 ttcgggttcgggttcacatcagcaaaaacggctcatcctcctgaccaaataccacgcgttc 432
DB 516 TTGGATCGGGTTTTTTTATCTCGTCTGATGTTATTTGTGACCAATAATCATGTGATT 575

QY 433 gccgggtatgggcagatcaaaagtcctcgtcaacgcaacgcaagcgcaataaccgcaaaactc 492
DB 576 TCTGATGGCACAAGTTACGCTGTGTTCTTGATGACGGTACAGAACTGAATGCAAAACTC 635

QY 493 atcgggttcggatgccaatccgatgcgccttcctgaaaaatcgacgcaacggaagagcta 552
DB 636 ATTGAACGGACCCACGACACTGATCTTCAGAGTATTAAAGTCAATGCAAAAAGAAAATTT 695

QY 553 cccgttcgtaaaatcggcaatcccaaaaatttgaacccggcggaatgggtcgctgccatc 612
DB 696 TCGTACGTTGATTTGGTGATGATTCAAAACTTCGTTGGTGATTTGGGTTTGTGCTATT 755

QY 613 ggcgcgccttcgggttcgacaaagtcgacgctgcgccttcgaaaaatcgacgcaacggaagcaga 672
DB 756 GGTAATCCATTTGGTCTTGGTGAACCTGACAGAGGTATCGTTTCAGCACGTGGACGT 815

QY 673 agcctgcccacgaagctcacacaccttcacccaacccgagcttgcccatcaatccgggc 732
DB 816 GATATCGGTACCGGTTTATGATGATTTTATTCAGATTGATGCTGCAGTTAATCGAGGA 875

QY 733 aattccggcgccgcgtgttcaacttaaaagacaggttcgctgcgccttcgaatcgcaaaata 792
DB 876 AATTCTGGAGGTCCAACCTTTTGTATCTTAACGGAAAGGTTTGGAGTGAATACGGCAATT 935
```

QY 793 tacagccgacgagcgattcatgggcatctctcttgccatcccgattgacgttgcctatg 852
Db 936 TTTTCTCTCTCTGGGGGCAAGCTTGGGATGCTTTTCGTATTCCGCGCAGCAACAGCGAAC 995
QY 853 aatgcgcgacagctgaaacacccgcaaaagtcacacgagcgagacaaactggcgctgatt 912
Db 996 GAGGTGTGCAACAACTTATCGAAAGAGTTAGTTCACGCTGGTGGCTTGGGGTTCAG 1055
QY 913 attcagaagatctactacggtttggcacagtcgttcctgctgataaaagccgagcgca 972
Db 1056 ATTGAGCTGTAAACAAAGAAATTTCTGATTCATCAATTTGTTTGAAGGAGGCTAAAGGTGCG 1115
QY 973 ttgattgcaaaatctctcccgagcccgagcccgagcaagtcgcccgcgagcgagcgagc 1032
Db 1116 TTAATTAACCGATCCATTAAGAGGGGCGACGCGCAAAA--GCTGGTATCAAGGCGAGTGAT 1172
QY 1033 atcgtctccagctcgcgagcgagaaatacgttctcccgagcacttcccgteatggtc 1092
Db 1173 GTTATTAATTCGGTAAATGGTGAGAGATTAATGATGTCGGTGATCAGCAAGCGTATT 1232
QY 1093 ggcgccattacgcccgggaaagagtcagctcgcgctgagcgcaaaagcgcaaaatc 1152
Db 1233 GCAATATGAGCCCGGAGAAACAGTAACTTTAGGAGTTTGGAAATCTGGTAAAGAGAG 1292
QY 1153 acaatcaagccaagctgg 1171
Db 1293 AATATTAAGGTAAACTTG 1311

RESULT 7

US-08-278-091-1
; Sequence 1, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-278-091-1

Query Match 10.0%; Score 139.4; DB 1; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;
QY 382 ggccttcatacatcagcaaaagcgctacatctctgacaaataccccacgctcgttcccggtatg 441
Db 1108 GTCATTTAATGCAAGCAAGGCTATGTTTAAACCAATATCATTTATTGATGAAGCT 1167
QY 442 ggcagtatcaaatcctgctcaacgacgagcgagataataccgcaaacatccatcggttcg 501
Db 1168 GATAAAATTAACCGTGAATTTACAGATGGCGTGAAATTTAAAGCAAAATTTAGTGGGTAAA 1227
QY 502 gatgtccaatccatgctcgcctcttgaataatcgacgcaaacggaagagctaccctgctgc 561
Db 1228 GATGAACACTATCAGATATTGCATTAGTACAGCTTGAAAAACCAAGTAATTTAAACAGAAATC 1287
QY 562 aaaaatcggaatccccaaaaatttgaaacccggcggaatggctcgtcgcacatcgcgccgcc 621
Db 1288 AAATTTGCTGATTCGACAAATTTACGCTAGGCGATTTTCACTGTTGCAATCGGTAATCCA 1347
QY 622 ttcggctttgacacagcgtgacgcgcgcatcgtctccgccaagg-----cagaagc 675
Db 1348 TTTGGTTTAGGTCAAACTGTGACATCAGGTATTGTTTCTGCATTGGTCTGCTTCAACAGT 1407
QY 676 ctgcccacgaaagctacacaccccttcacccaaacgagcgttgcataatcccgggcaat 735
Db 1408 TCTGACAGTGGCACTTATGAAAACTATATTCAAAACCGATCAGCAGTAAACCCGGGTAAT 1467
QY 736 tccggcgcccgctgttcaacttaaaagacaggtcgtcgggcacatccatcccggaat 795
Db 1468 TCGGGTGGAGCGTTAGTAACTTAAATGGCGAACTTATTGGAATTAATACCGCAATATT 1527
QY 796 agccgcagcgcggaattcatggcctctcttggcctccatcccgattgacgttgcataaat 855
Db 1528 TCTCCAAAGCGGTGGCAATCGAGGAATTCGCTTTGCCATTTCCAAGTAATCAAGCAAGCAAT 1587
QY 856 gtcgcccgaacagctgaaaaacacccggcaaaagtcacacgagcgagcaactggcgctgattatt 915
Db 1588 TTAGTCAACAAATTTTAGAATTTGGTCAAGTGCCTCGCGGATTTGCTTGTGTTTAAAGGT 1647
QY 916 caggaagatctactacggtttggcacagtcgttcggctctggataaaagccagcgcgatg 975
Db 1648 GCGCAACTCAATGCTGATTTAGCCAAAGCGCTTTAATGTAAAGCGCAACAAAGCGCATTT 1707
QY 976 attgccaataatctctcccgccagcccgagcaacgctgcggcgctcagcgcgcgacatc 1035
Db 1708 GTAAGTGAAGTTTACCGAAATCTGCTGTGTAAGAAAGCAGGACTTAAAGCGGGCGATATT 1767
QY 1036 gtctcagcctcgacggcgagaaatacgttctctccggcgacattcccgctcagtggtcgcc 1095
Db 1768 ATCAGCGCGATGAACGGTCAAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAAATCGCA 1827
QY 1096 gccattacccgggaaagaaagtcacccctcggtatggcgcaagcgcaa 1146
Db 1828 ACCACTGGTGCAGGCAAGAGATTAGCTTACTTACTTACTTACTTACTTACTTACTTACTTACT 1878

RESULT 8

US-08-483-859-1
; Sequence 1, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

QY	856	gtcgccgaacagctgaaacacacccggaacagcgagacaactcgggcgctgatt	915
Db	1588	TTAGTGCACAAATTTTGAATTTGGTCAAGTGGCTGCGGATGCTTGGTATTAAAGGT	1647
QY	916	cagggaagtatctcagtttgggcacagctcgcttcggtctgataaagccagcgcgcatgg	975
Db	1648	GGCGAACTCAATGCTGATTTAGCCAAAGCCTTTATGTAAAGCGGCACAAAGGCGCATTT	1707
QY	976	atggccaaaactcttcgccggcagcccccagaaacgctgcggcgctcgaggcgcgacatc	1035
Db	1708	GTAAGTGANGTTTACCGAAATCTGCTGCTGAAAAGCAGGACTTAAAGCGGCGATATT	1767
QY	1036	gtcctcagctcagcagcgggagaaataagcttcttcggcgaccttcgcgtcatggtcgc	1095
Db	1768	ATCACGGCGATGACGGTCAAAAATCTCAAGTTTCGCTGAANTCTGTCGCAAAAATCGCA	1827
QY	1096	gccattacgcccgggaaagaagctcagccttcggctatggcgcaagcgaa	1146
Db	1828	AGCAGCCCTCAGGCAACAGAGATTAGCTTGACTTACTTACGTGATGGCAAA	1878

RESULT 9
US-08-472-173-1
; Sequence 1, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-ping
; APPLICANT: CHONG, Pele
; APPLICANT: OMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sim & McBurney
; ADDRESS: Suite 701 330 University Avenue

	Query Match	10.0%;	Score 139.4;	DB 1;	Length 2894;
	Best Local Similarity	49.8%;	Pred. No. 2.3e-29;		
	Matches 384;	Conservative	0; Mismatches 381;	Indels	Gaps
QY	382	ggcttcacatcagcaaaaacggtcgatacctctgaccaatacccaacgcgtcttgcgggtatg	441		
Db	1108	GTCATTATTAATGCAAGCAAAGGCTATGTTTAAACCAATAATCATGTATTATGATGAAGCT	1167		
QY	442	ggcagttatcaaagtctctgctcaacgacaacgagggaatatatacccacaaactcatcggttcg	501		
Db	1168	GATAAANAATTACC CGTGCAATTACAAGATGGCGGTGAATTTAACGCAANAATTAGTGGGTAAA	1227		
QY	502	gatgtccaaatccgatgtcgccctctctgaaaaatcgacgacaaacggaagagtgaccocgtcgtc	561		
Db	1228	GATGAACATATCAGATATATGCATTAGTACAGCTTGAAAAACCAAGTAGTAATTTAACAGAAATC	1287		
QY	562	aaaaatcgccaattccccaaaaatttgaaacccggggcgaatgggtcgtcgccatcgcgcgccgc	621		
Db	1288	AAATTTGCTGATPCCGACAAATTAGCGTAGCGGATTTCACTGTGCAATCGGTATATCCA	1347		
QY	622	ttagcgtttgacaacagcgtgacgcgcgcgcgtctgtccgccaaagg-----cagaagc	675		
Db	1348	TTTGGTTTTAGGTCAAACTGTGACATCAGGTATTTGTTCTGATTGGGTCGTTCAACAGGT	1407		
QY	676	ctgcccacaacgaagctacacaccttcatccaaacgagcttgccatcaatccgggcaat	735		
Db	1408	TCTGACAGTGGCACATTAGAAACATATATTCAACCGGATGAGCAGTAAACCCGGGTAAAT	1467		
QY	736	tccggcgccgcgctgttccaacttaaaaggacaggtcgctcggcacatattgcgcaaatatc	795		
Db	1468	TCGGGTGGAGCGTTAGTAAACITAAATGCGCAACTTATTGGAATTAATACCGCAATTAAT	1527		
QY	796	agccgcagcggcgattcatagggcatctcctttggcattcccgattgacgttgccatgaat	855		
Db	1528	TCGCCAAGCGGTGGCAATCAGGAATPTGCTTTGGGATTCGAATTAATCAAGCAAGCAAT	1587		

```
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggttcacatcagcaaaacgggtacatccgaccataccacgcgtgtgcccgtatg 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 GTCAATTAATGAAGCAAGAGGTATGTTTAAACCAATATCATGTTATTGATGAAGCT 1167

QY 442 ggcagtatcaaaagtcctgctcaacagcaagcgcaataatccgcaaaactcatcggttcg 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 GATAAAATTAACCGTGCATTAACAAGATGGCGGTGAATTTAAAGCAAAATTAAGTGGGTAAA 1227

QY 502 gatgtcaatcccgatcgccctctgaatacgcacgcaacggaagagctaccgctgctc 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GATGAACATATCAGATATGATTAAGTACAGCTTGAACCAACCAAGTAATTTAACAGAAATC 1287

QY 562 aaatgggcaatccccaaatattgaacacggcggaatgggtcgtgccatcgggcgccccc 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 AAATTTGCTGATTCGACAAATTAACGGGTAGCGGATTTCACTGTTGCAATCGGTAATCCA 1347

QY 622 ttcgggtttgacaacagcggtgacccgcgcatcggtcgtccgcaaaagg-----cagaagc 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 TTTGGTTTAGGTCAAACTGTGACATCAGGTATTTGTTCTGCATTTGGGTTCGAACAGGT 1407

QY 676 ctgcccacgaaagctacacaccccttcacaaacgacgttgccatcaatcccggaat 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 TCTGACAGTGGCACATTTGAAAACATATATTCAAACCGATGCAGCAGTAACCCGCGGTAAT 1467

QY 736 tccggcgcccgctgttcaacttaaaagacaggtgctcgccatcaatccgcaaatatc 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 TCGGGTGGAGCGCTTAGTAACCTTAAATGCGAATCTATTGGAATTAATACCGCAATATT 1527

QY 796 agccgcagcgcggtatcatgggcatctcttgcaccccgatcgacgttgccatgaat 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1528 TCTCCAAGCGGTGGCAATTCAGGAATTTGCTTTTGGGATTTCCAAAGTAATCAAGCAAGCAAT 1587

QY 856 gtcgcccgaacagctgaaaaacacccgcaaaagtcacacgacgacactggcggtgattatt 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1588 TAGTGCACCAAAATTTTGAATTTTGGTCAAGTGGCTCGCGGATTTGCTTGGTATTAAAGGT 1647

QY 916 caggaaagatccctacggttttgacagtcgctcgtggttgataaagcagcgcgcatg 975
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1648 GCGGAACCTCAATGCTGATTTAGCCAAAAGCCCTTTAATGTAAAGCGGCAACAAGCGCAATT 1707

QY 976 atgcccataatcttccggcagcccgacgaacgtgcccgcctgcagcgcgggcgacatc 1035
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1708 GTAAGTGAAGTTTTTACCAGAAATCTGCTGCTGAAAAGACGAGACTTAAAGCGGCGGATATT 1767

QY 1036 gtccctcagcctcgacgaggaataacgttcttccggcgacccctcccgctcatggtcggc 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1768 ATCAGCGGATGAACGCTCAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAAATCGCA 1827

QY 1096 gccattacgcccggaaagaagtcagcctcggtggtatggcgcaaaagcgaa 1146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1828 ACCACTGGTGCAGCAAGAGATTAGTGTGACTTACTTACGTGATGGCAAA 1878

RESULT 10
US-08-487-167-1
; Sequence 1, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-167-1

Query Match 10.0%; Score 139.4; DB 3; Length 2894;
Best Local Similarity 49.8%; Pred. No. 2.3e-29;
Matches 384; Conservative 0; Mismatches 381; Indels 6; Gaps 1;

QY 382 ggttcacatcagcaaaacgggtacatccgaccataccacgcgtgtgcccgtatg 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 GTCAATTAATGAAGCAAGAGGTATGTTTAAACCAATATCATGTTATTGATGAAGCT 1167

QY 442 ggcagtatcaaaagtcctgctcaacagcaagcgcaataatccgcaaaactcatcggttcg 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 GATAAAATTAACCGTGCATTAACAAGATGGCGGTGAATTTAAAGCAAAATTAAGTGGGTAAA 1227

QY 502 gatgtccaatccgagtcgcccctctgaaatcgacgcaacggaagagctaccgctgctc 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 GATGAACATATCAGATATGATTAAGTACAGCTTGAACCAACCAAGTAATTTAACAGAAATC 1287

QY 562 aaatcgcgcaatccccaaatattgaacacggcggaatgggtcgtgccatcgggcgccccc 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 AAATTTGCTGATTCGACAAATTAACGGGTAGCGGATTTCACTGTTGCAATCGGTAATCCA 1347

QY 622 ttcgggtttgacaacagcggtgacccgcgcatcggtcgtccgcaaaagg-----cagaagc 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 TTTGGTTTAGGTCAAACTGTGACATCAGGTATTTGTTCTGCATTTGGGTTCGAACAGGT 1407

QY 676 ctgcccacgaaagctacacaccccttcacaaacgacgttgccatcaatcccggaat 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 TCTGACAGTGGCACATTTGAAAACATATATTCAAACCGATGCAGCAGTAACCCGCGGTAAT 1467

QY 736 tccggcgcccgctgttcaacttaaaagacaggtgctcgccatcaatccgcaaatatc 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 TCGGGTGGAGCGCTTAGTAACCTTAAATGCGAATCTATTGGAATTAATACCGCAATATT 1527

QY 796 agccgcagcgcggtatcatgggcatctcttgcaccccgatcgacgttgccatgaat 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1528 TCTCCAAGCGGTGGCAATTCAGGAATTTGCTTTTGGGATTTCCAAAGTAATCAAGCAAGCAAT 1587

QY 856 gtcgcccgaacagctgaaaaacacccgcaaaagtcacacgacgacactggcggtgattatt 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1588 TAGTGCACCAAAATTTTGAATTTTGGTCAAGTGGCTCGCGGATTTGCTTGGTATTAAAGGT 1647
```



```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,149
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-296-149-1

```

	Query Match	10.0%	Score 139.4	DB 4	Length 2894
	Best Local Similarity	49.8%	Pred. No. 2.3e-29		
	Matches 384	Conservative 0	Mismatches 381	Indels 6	Gaps 1
QY	382	ggcttcacatcagcaaaaacggctacatcctgaccatacccaacgtcggtgcccggatg	441		
Db	1108	GTCATTATTAAATGCAAGCAAAAGGCTATGTTTAAACCAATAATCATGTATTATTGATGAAGCT	1167		
QY	442	ggcagatcaaaagtcctgctaacgcagacaagcgcgaaataacgcgcaaaactcatcggtcgg	501		
Db	1168	GATAAAATTACCGTGCAATTACAGATGGGCGTGAAATTTAAAGCAAAATTTAGTGGGTAAA	1227		
QY	502	gatgtccaaatccgatgtgcgctcttgaaaatcgacgcgaacggaagatcaccgcgtc	561		
Db	1228	GATGAACATATCAGATATTGCATTAGTACAGCTTGAAAAACCAAGTAATTTAACACAGAAATC	1287		
QY	562	aaatcggcaatcccaaaaatttgaaacccggcggaattgggtcgctgcacatcgcgcgccc	621		
Db	1288	AAATTTGCTGATTCCGACAAATTAAGCGTAGGCGGATTTCACTGTTGCAATCGGTAAATCCA	1347		
QY	622	tccggcttgcacacagcgtgacgcgcgcgcgtcgtcgccgcaaaagg-----cagaagc	675		
Db	1348	TTTGCTTTAGCTCAAACTGTGCATCAGTATTGTTCTGCTATGGGTGCGTTCAACAGGT	1407		
QY	676	ctgcaccaacgaagctacacaccccttcacaaaccgacgttgccatcaatccgggcaat	735		
Db	1408	TCTGACAGTGGCACTTATGAAACATATATTCAAACCGATGCAGCAGTAACCGCGGTAAAT	1467		
QY	736	tccggcgccgcgtgttcaactttaaagacaggtcgctcggcatacaattcgcaaatatac	795		
Db	1468	TCGGGTGGAGCGTTAGTAAACTTAAATGGCAACTTATTGGAATTTAATCCGCAATATT	1527		
QY	796	agcgcagcggcggtatcatcggtcctcttgcacatcccgattgacgttgcacatgaat	855		
Db	1528	TCTCCAGCGGTGGCAATGCAGGAATTCCTTTGGGATTCGAAGTAATCAAGCAAGCAAT	1587		
QY	856	gtcccgcaacagctgaaaaacacccggcaaaagtcccaacgcgacaaactggcggtgattatt	915		
Db	1588	TTAGTGAACAATAATTTAGAAATTTGGTCAAGTGCCTCGCGATTGCTTGGTATTAAAGGT	1647		
QY	916	caggaaagtatcctcaggttttgacacagtcgcttcgggtctggataaagccagcgggcattg	975		
Db	1648	GGCGAACTCAATGCTGATTTTAGCCAAAGCCCTTTAATGTAAAGCGCGCAACAAGCGCATTT	1707		
QY	976	attgccaaaaatcccttcccgcgagcccgcaagaagtcggcgccctgcagcggcggaacatc	1035		
Db	1708	GTAAGTGAGTTTACCGAAATCTGCTGCTGAAAAAGCAGGACTTTAAAGCGGGCGATATT	1767		
QY	1036	gtctcagctcgcagcggcggaataacgtttcttcggcgacacttccctcatgtatgcgc	1095		

```

Db 1768 ATCAGCGGATGAAGCGTCAAAAAATCTCAAGTTTCGCTGAAATTCGTGCAAAATCGCA 1827
QY 1096 gccattagcgggaaaaaagtacgctcgcgcgtatgcgcaaaagcgaa 1146
Db 1828 ACCATGGTGACAGCAAGAGATTACTTGACTTACTTACGTATGCGAAA 1878

RESULT 13
US-08-801-499-1
; Sequence 1, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/801,499
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/POCKET NUMBER: 1038-671 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-801-499-1

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: April 14, 2000, 15:42:07 ; Search time 1307.83 Seconds
(without alignments)
4027.321 Million cell updates/sec

Title: US-09-388-090-3
Perfect score: 1395
Sequence: 1 gtgtcaaaaataccaata.....tcagggaagtcocccgtcaa 1395

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues 9077268

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100.6	7.2	490	34	AA479844	AA479844 zu43h07.f
2	97.4	7.0	489	28	AA110074	AA110074 mo51b12.f


```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 489)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE        The WashU-HHMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1405092.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:37879
Seq primer: -28M13 rev1 from Amersham
High quality sequence stop: 446.
FEATURES
source
1..489
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:557087"
/clone_lib="Life Tech mouse embryo 10 5dpc 10665016"
/tissue_type="embryo"
/dev_stage="10.5dpc embryos"
/lab_host="DH10B"
/notes="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. 10.5dpc embryos, pCMV-SPORT2 vector."
BASE COUNT   130 a 116 c 140 g 103 t
ORIGIN
Query Match      7.0%; Score 97.4; DB 28; Length 489;
Best Local Similarity 53.8%; Pred. No. 1.2e-14;
Matches 231; Conservative 0; Mismatches 186; Indels 12; Gaps 1;

QY 366 attgaacttcggttcggttcattcatcagcaaaacggctacatcctgaccaataccca 425
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 AGTGGCCAGTGGGTCAAGGATTTCATCGTATCGGAGGATGGACTGATTGTGACAAATGCTCA 105
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 426 cgtcgttgccggtatgggcagtagtcaaaagtcctgctcaacgacagcggaataaccgc 485
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 CBTGGTCAACCAAAACCGGGTCAAGGTGAGCTGAAGATGGAGCTACCTATGAAGC 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 486 caaactcatcgttcggtatcccaatccgatgcgcctctgaaatctgaaatcgacgcaacgga 545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CAAATCAAGGATCTGGATGAAAGGCGGACATTCGCTTATCAGATTTGACCAACGAGG 225
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 546 agagctaccgctgcacaaatcggaatcccaaaatttgaacccggcggaatgggtcgc 605
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 AAAGCTGCCAGTCTCTGCTGTCGGCCCTCTCTCAGAGCTGAGACCTGGAGATTTGTAGT 285
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 606 tgccatcgcgcgccttcggtttgacacagcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 TGCCATTGGAAGCCCTCTTTCTCTTCAAAACACAGTCAAGTCCGATCGTTCAGCACACAC 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 666 aggcagaagc-----ctgcccacgaagtagtacacaccccttcacccaaccca 713
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 CACAGCAGGCGGCAAGAGCTGGGACATTCGGAATCCGATATGACATACATTCAGACAGA 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 714 cgttgccataatccgggcaattccggcgccgctgttcaacttaaaagacagcgtgcgt 773
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 CGCTATCATCAATATGGAGATTCCGGAGGCCCGTTAGTAAACCTGGATGCCGAGGTGAT 465
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 774 cggcatcaa 782
||| ||| |||
Db 466 TGGGATTAA 474

RESULT 3
LOCUS    A1159029      456 bp      mRNA      EST      02-OCT-1998
DEFINITION v283a08.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
IMAGE:133046 5' similar to TR:Q92743 Q92743 NOVEL SERINE PROTEASE.
; mRNA sequence.
ACCESSION A1159029
VERSION    A1159029.1 GI:3692211
KEYWORDS   EST.
SOURCE     Mouse mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 456)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE     The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2286627.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:692590
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 241.
FEATURES
Location/Qualifiers
1..456
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:133046"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTACCAATCTGAATGGAGCGGAGCGGCCGGAATGGTGTGTGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT   124 a 107 c 127 g 97 t 1 others
ORIGIN
Query Match      6.8%; Score 95.4; DB 42; Length 456;
Best Local Similarity 53.9%; Pred. No. 3.7e-14;
Matches 226; Conservative 0; Mismatches 181; Indels 12; Gaps 1;

QY 376 ggttcgggttcattcatcagcaaaacggctacatcctgacccaatccacgcgtgtgcc 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 GGGTCAGGATTTCATCGTATCGGAGGATGGACTGATGTGACAAATGCTCAGTGGTGCACC 69
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 436 ggtatgggcagtagtcaaaagtcctctcaagacagcagcgaataatataccgcgaactcatc 495
 Db 70 AACAAAAACCGGTCTCAGGTGAGCTGAAGAATGAGGACTACTATGAGCCCAAAATCAAG 129
 QY 496 ggtcggatccaatccgctctgctcctctgaaatcgaacgcaagagagctaccc 555
 Db 130 GATGTGGATGAAGAAGCGGACATTCGCTTATCAGATTTGACACCGGGAAGCTGCCA 189
 QY 556 gtcgtcaaaatcgcaatcccaaaattgaaacccggcggaatggctgcgtgcccattcgc 615
 Db 190 GTCCTGCTGTCGGCCCTCTCAGAGCTGAGACCTGGAGATTTGATGTGCCATTTGA 249
 QY 616 gcgccttcggcttggaacagcagtcgacccgcccgcctctgctgcgcgaaggcagaagc 675
 Db 250 AGCCCTCTTTCTCTCAAAACACAGTACCACCTGGATCGTCAGCACCCACCGCAGGC 309
 QY 676 -----ctgcccacgaagctcacaccccttcaaccccttcaacccgacgttgcctc 723
 Db 310 GGCAGAGCTGGGACTTCGGAACTCCGATATGAGCTATCATTCAGACAGCGCTATCATC 369
 QY 724 aatccgggcaatccggcgccgctgttcaacttaaaagacagcaggtcgtcggcatcaa 782
 Db 370 AATTATGGAATCCGGAGCCGCTTAGTAAACCTGGATGCGGAGGTGATTGGGATTAA 428

RESULT 4
 W17262 474 bp mRNA EST 29-APR-1996
 LOCUS zbl17a08.r1 Soares fetal_lung_NbHL19w Homo sapiens cDNA clone
 DEFINITION IMAGE:302294 5' similar to SW:HTRA_SALTY P26982 PROTEASE DO
 PRECURSOR ;, mRNA sequence.
 ACCESSION W17262
 VERSION W17262.1 GI:1291697
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 474)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT On May 18, 1995 this sequence version replaced gi:811473.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: ESTPrimer
 High quality sequence stop: 446.
 Location/Qualifiers
 1. .474
 /organism="Homo sapiens"
 /db_xref="GDB:1247226"
 /db_xref="taxon:9606"
 /clone="IMAGE:302294"
 /clone.lib="Soares_fetal_lung_NbHL19w"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: lung; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAGTGGGAGCGGCGCAATTTTTTTTTTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Fatima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NbHL19w."

BASE COUNT 122 a 121 c 135 g 93 t 3 others
 ORIGIN

Query Match 6.6%; Score 92.2; DB 25; Length 474;
 Best Local Similarity 57.4%; Pred. No. 2.5e-13;
 Matches 166; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 376 ggtcggcttcctcatcagcaaaacggctacatcctgaccaataccccacgtctgtgc 435
 Db 41 GGGTCTGGTATTATTGTGCGGAAGATGGACTGATCGTGACAAATGCCACGCTGGTGACC 100
 QY 436 ggtatgggcagtagtcaaaagtcctgctcaacgacagcgaataatataccgcgaactcatc 495
 Db 101 AACAAAGCACCGGGTCAAAAGTTGAGCTGAAGAACGGTGCACATTACGAAGCCAAAATCAAG 160
 QY 496 ggtcggatgctccaatccgctgctgcctcttgaaatcgacgcaacggaagagctaccc 555
 Db 161 GATGTGGATGAGAAGACAGACATCGCACATCATCAAAATTGACCACCGGCAAGCTGCCCT 220
 QY 556 gtcgtcaaaatcggaatcccaaaatttgaaacggcggaatggctgcgtgcctcgcctgcgc 615
 Db 221 GTCTGCTGCTTGGCGCTCTCTCAGAGCTGCGCGCGGAGAGTTCTGTTGCGCATCGGA 280
 QY 616 gcgccttcggcttgacaacagcgtgacccgctgacccgctgctgctgcgcga 664
 Db 281 AGCCCGTTTTTCCCTTCAAAACACAGTCACCCCGGATCGTGAGCACCA 329

RESULT 5
 R52327 485 bp mRNA EST 18-MAY-1995
 LOCUS YG75a09.r1 Soares infant brain INTB Homo sapiens cDNA clone
 DEFINITION IMAGE:38967 5' similar to SP:HTRA_SALTY P26982 PROTEASE DO
 PRECURSOR ;, mRNA sequence.
 ACCESSION R52327
 VERSION R52327.1 GI:814229
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 485)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1476
 High quality sequence stops: 391 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1476 Std Error: 0.00
 Seq primer: W13RP1
 High quality sequence stop: 391.
 Location/Qualifiers
 1. .485
 /organism="Homo sapiens"
 /db_xref="GDB:411508"

FEATURES
 source

```

/db_xref="taxon:9606"
/clone="IMAGE:38967"
/clone_lib="Soares infant brain lNIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="PH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site.1: Not
I; Site.2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
ACTGGAAGATTCGGCGGCGAGGAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      126 a  121 c  135 g  103 t
ORIGIN

```

```

Query Match      6.6%; Score 92.2; DB 22; Length 485;
Best Local Similarity 57.4%; Pred. No. 2.5e-13;
Matches 166; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 376 ggttcgggttcacatcagcaaaacggctacatcctcgaccataaccacgtcgttgc 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 GGGTCGTGGTTATTGTGTCGGAAGATGGACTGATCGTGACAAATGCCACGGTGTGACC 64

QY 436 ggtatgggagtcataaagtcctgtcaacgacgaagcgaatataccgcgaactcacc 495
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AACAAACGACCGGGTCAAAAGTTGAGTGAAGAACGGTGCCACTTACGAAGCCAAAATCAAG 124

QY 496 gtttcgagtcacaatccgagtcgccttcctgaaacacgcaagcgaagcgtacc 555
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 GATGTGGATGAGAAAGCAGACATCGCACTCATCAAAATGACCACCGGCAAGCTGCCT 184

QY 556 gtcgtcaaatcgcaatcccaaaatttgaaacggcgcaatggctgcgtgccatcggc 615
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 GTCTCTGCTGTGGCGCTCCTCAGAGCTGCGCGCGGAGAGTCTGTGGTGGCCATCGGA 244

QY 616 ggcgccttcggttgacacacgctgacccgcgccatcgtgtccgcga 664
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 AGCCCGTTTTTCCCTTCAAAACACAGTCACACCGGGGATCGTGAGCACCA 293

```

```

RESULT 6
F07134
LOCUS      F07134      350 bp      mRNA      EST      20-FEB-1995
DEFINITION HSCIWG081 normalized infant brain cDNA Homo sapiens CDNA clone
c-1wg081, mRNA sequence.
ACCESSION F07134
VERSION    F07134.1 GI:672779
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 350)
AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
On Sep 21, 1992 this sequence version replaced gi:279153.
Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'interationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

```

```

Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-1wg08
Insert Length: 792 Std Error: 0.00
Seq primer: (-21)M13_universal
High quality sequence stop: 371.
FEATURES             Location/Qualifiers
     source            1..350
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: Lfamid BA; Site.1: HindIII;
                     Site.2: NotI; sex:Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient; tissue_type=total
                     brain; total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the HindIII -> NotI sites of the
                     Lfamid BA vector. Clone library from B.Souares, Psychiatr;
                     Dept. Columbia University, USA. Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT      88 a   95 c  100 g   66 t
ORIGIN

```

```

Query Match      6.4%; Score 89.6; DB 21; Length 350;
Best Local Similarity 56.7%; Pred. No. 9.8e-13;
Matches 164; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 376 ggttcgggttcacatcagcaaaacggctacatcctcgaccataaccacgtcgttgc 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 GGGTCGTGGTTATTGTGTCGGAAGATGGTCTGATCGTGACAAATGCCACGGTGTGACC 65

QY 436 ggtatgggagtcataaagtcctgtcaacgacgaagcgaatataccgcgaactcacc 495
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 AACAAACGACCGGGTCAAAAGTTGAGTGAAGAACGGTGCCACTTACGAAGCCAAAATCAAG 125

QY 496 gtttcgagtcacaatccgagtcgccttcctgaaacacgcaagcgaagcgtacc 555
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 GATGTGGATGAGAAAGCAGACATCGCACTCATCAAAATGACCACCGGCAAGCTGCCT 185

QY 556 gtcgtcaaatcgcaatcccaaaatttgaaacggcgcaatggctgcgtgccatcggc 615
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 GTCTCTGCTGTGGCGCTCCTCAGAGCTGCGCGCGGAGAGTCTGTGTGCGCATCGGA 245

QY 616 ggcgccttcggttgacacacgctgacccgcgccatcgtgtccgcga 664
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 AGCCCGTTTTTCCCTTCAAAACACAGTCACACCGGGGATCGTGAGCACCA 294

```

```

RESULT 7
AI903651/c
LOCUS      AI903651      337 bp      mRNA      EST      01-DEC-1999
DEFINITION QY-BT032-190299-136 BT032 Homo sapiens cDNA, mRNA sequence.
ACCESSION AI903651
VERSION    AI903651.1 GI:6494038
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
TITLE     The FAPESP/LICR Human Cancer Genome Project
JOURNAL   Unpublished (1999)
COMMENT   On Dec 20, 1995 this sequence version replaced gi:1134214.
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```


RESULT 10
 C18770 LOCUS 346 bp mRNA EST 02-OCT-1996
 DEFINITION C18770 Human placenta cDNA (TFujiwara) Homo sapiens cDNA clone
 GEN:567E01 5', mRNA sequence.
 C18770
 C18770.1 GI:1580372
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 346)
 AUTHORS Fujiwara,T., Hirano,H., Hishigaki,H., Horie,M., Kawai,A., Kuga,Y.,
 Kyushiki,H., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y.,
 Shinomiya,H., Suzuki,M., Takaichi,A., Takeda,S., Watanabe,T.,
 Maekawa,H., Nakamura,Y. and Takahashi,E.
 TITLE Otsuka cDNA project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393740.
 Contact: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 4-3-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.
 FEATURES
 source Location/Qualifiers
 1..346
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-567E01"
 /clone_lib="Human placenta cDNA (TFujiwara)"
 /tissue_type="placenta"
 BASE COUNT 32 a 90 c 77 t
 ORIGIN
 Query Match 5.8%; Score 81.2; DB 28; Length 346;
 Best Local Similarity 56.3%; Pred. No. 1.3e-10;
 Matches 152; Conservative 0; Mismatches 118; Indels 0; Gaps 0
 QY 376 ggttcgggttcacatcagcaaaaacggctacatcctgacacataaccacagctgtgcc 435
 Db 77 GGGTCGGGTTATGTGTCGGAAGACTGATCGTGACAAATGCCACGTGTGTGACC 136
 QY 436 ggtatgggcagtcatacaagtcctgctcaacagacagcgcgaatataaccgcgaactcacc 495
 Db 137 AACAAAGCACCGGGTCAAAAGTTGAGCTGGAAGAACGGTGGCCACTTACGAAGCCAAAATCAAG 196
 QY 496 ggttcggatgctcaatccgatgtccctctgaaatcgacgcgaacgagagctacc 555
 Db 197 GATGTGGATGAGAAAGACAGACATCCGACTCATCAAAATGACCACGGGCAAGCTGCCT 256
 QY 556 gtctcgaataatcggaatcccaaaatttgaacacggcggaatgggtcgctccatcgcc 615
 Db 257 GTCTGCTGCTTGCGCGTCTCAGAGCTGCGCGGGGAGAGTTGCTGTGTCGCCATCGGA 316
 QY 616 ggcgccttggcgtttgacaaacagcgtgacc 645
 Db 317 AGCCCGGTTTTCCTTCAAAACACAGTCACC 346
 RESULT 11
 AA644825 LOCUS 375 bp mRNA EST 28-OCT-1997
 DEFINITION AA644825 Barstead mouse myotubes MFLR5 Mus musculus cDNA clone
 IMAGE:1152856 5', similar to SW:HTRA_SALTY P26982 PROTEASE DO
 PRECURSOR ;, mRNA sequence.
 ACCESSION AA644825
 VERSION AA644825.1 GI:2571254
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus


```

AI982113      623 bp      mRNA      EST      15-SEP-1999
LOCUS      pat.pk0074.e8.f chicken activated T cell cDNA Gallus gallus cDNA
DEFINITION      clone pat.pk0074.e8.f 5' similar to serine protease, mRNA sequence.
ACCESSION      AI982113
VERSION      AI982113.1 GI:5885141
KEYWORDS      EST.
SOURCE      chicken.
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 623)
Tirunegaru,V.G., Sofer,L. and Burnside,J.
An expressed-sequence-tag database of activated chicken T cells:
sequence analysis of 5596 clones
Unpublished (1999)
On May 20, 1999 this sequence version replaced gi:4878134.
Contact: Joan Burnside
Molecular Endocrinology
University of Delaware
40 Townsend Hall, Newark, DE 19717, USA
Tel: 302 831-1345
Fax: 302 831-3651
Email: joan@udel.edu
Seq primer: 27.
FEATURES             Location/Qualifiers
     source
     1..623
     /organism="Gallus gallus"
     /db_xref="taxon:9031"
     /clone="pat.pk0074.e8.f"
     /clone_lib="Chicken activated T cell cDNA"
     /sex="male"
     /cell_type="Con A-activated splenic T cell"
     /lab_host="E.coli TOP10 F'"
     /note="Vector: pcDNA3"
BASE COUNT      187 a 137 c 156 g 131 t 12 others
ORIGIN
Query Match      5.6%; Score 77.8; DB 63; Length 623;
Best Local Similarity 54.3%; Pred. No. 1.2e-09;
Matches 157; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY      376 ggttcgggttcatactcgcgcaaaacggcgtacatcctgcacaaatccacgcgtgtgccc 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      67 GGGTCGGGTTTCATTGCTCTCGGAGGATGGCTGATAGTGACCAACGCTCACGTGGTCACC 126
QY      436 ggtatgggcagatcaaatcctcctgcacgcaaacgcaaatataccgcaaacatcacc 495
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      127 AACAAACAGCGGTGAAGGTGGAGCTGAAGATGGAGAACATACCAAGCTAAATATAA 186
QY      496 ggttcggatgccatccgatcgtgcctctgaaatcgacgcaacggaagagctacc 555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      187 GAGTTGATGAGAAAGCTGACATTCGACATAATTAATAGATGCTCAGGGTAATGGCCA 246
QY      556 gtctgcaaatcggcaatcccaaaatttgaaacccggcggaatgggtcgtcgtccatcggc 615
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      247 GTCTTCTGCTTGGTCACTGAGTGGAGACTTGAGCGCGGAGAGTTTGTGTTGGCATTTGG 306
QY      616 ggcgccttcggcttgacaacagcgtgacccgcgcgtcgtgctccacca 664
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      307 AATCCATTTTCCCTCCAAAACACAGTACACACGCGGGATCGTCAACACCA 355
RESULT      14
AA349615      292 bp      mRNA      EST      21-APR-1997
LOCUS      EST36620 Infant brain Homo sapiens cDNA 5' end similar to similar
DEFINITION      to Protease DO, mRNA sequence.
ACCESSION      AA349615
VERSION      AA349615.1 GI:2002068
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 292)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudet,D.M., Shirley,R.,
Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL      12140200
MEDLINE
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1394192.
Other_ESTs: THCI72755
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018899056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source
     1..292
     /organism="Homo sapiens"
     /db_xref="ATCC (inhost):150705"
     /db_xref="taxon:9606"
     /clone_lib="Infant brain"
     /sex="female"
     /dev_stage="infant"
     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
           Site_2: NotI"
BASE COUNT      76 a 88 c 79 g 47 t 2 others
ORIGIN
Query Match      5.0%; Score 69.6; DB 32; Length 292;
Best Local Similarity 57.0%; Pred. No. 1.1e-07;
Matches 126; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY      444 cagtatcaagtctgctcaacgacgacgacgcaataacgcgaactcgtgttcgga 503
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1 CCGGGTCAAGTTAGCTGAAGAACGGTGCCACTTACGAAGCCAAATCAAGGATGTGA 60
QY      504 tgtccaatccgatgtgcctctctgaaatcgacgcaacggaagagctaccgtctcaa 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      61 TGAGAAAGCAGACATCGCACATCAAAATTCACCACCGGGCAATGCCTGTCTCTGCT 120
QY      564 aatcggcgaatcccaaaatttgaaacccggcggaatgggtcgtcgtcgtcgtcgccct 623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      121 GCTTGCCCTCTCTCAGAGCTCGGGCCGGGAGAGNTCGTGGTCGCCATCGGAAGCCGTT 180
QY      624 cggcttgacaacagcgtgaccgcggcatcgtgttcgcca 664
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      181 TTCCCTTCAAAACACAGTCACCAACCGGGATCGTGAGCACCA 221

```


RESULT 15

AA186646 485 bp mRNA EST 10-MAR-1998
 zp64902.r1 stratagene endothelial cell 937223 Homo sapiens CDNA
 clone IMAGE:625010 5' similar to SW:HTRA_SALTY P26982 PROTEASE DO
 PRECURSOR ;, mRNA sequence.

ACCESSION AA186646
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 485)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
 Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WASHU-NCI human EST Project
 Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394135.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1515 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 388.

FEATURES
source

1..485
 /organism="Homo sapiens"
 /db_xref="GDB:5047164"
 /db_xref="taxon:9606"
 /clone="IMAGE:625010"
 /clone_lib="Stratagene endothelial cell 937223"
 /dev_stage="umbilical vein, 1 passage"
 /lab_host="SOLR (kanamycin resistant)"
 /notes="Vector: pBluescript SK-; Site_1: ECORI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dT.
 Umbilical vein endothelial cells, passaged once. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATCGGCACGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 136 a 129 c 124 g 95 t 1 others
 ORIGIN

Query Match 5.0%; Score 69.4; DB 29; Length 485;
 Best Local Similarity 55.5%; Pred. No. 1.5e-07;
 Matches 201; Conservative 0; Mismatches 147; Indels 14; Gaps 3;

QY 398 aaacggctacatcctgaccataaccacacgtgtggtgcggtatgggagatcaaatgcc 457
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2 AAGATGGACTGATCGTGACAAATGCCCGCTGGTGACCAACAGCCGGGTCAAAGTTG 61
 QY 458 tgcctaacgacagcggaatataccgccaaactcatcggttcggtatgtccaatccgatg 517
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 62 AGCTGAAGACGGTGCCTACTTACGAAGCAAAATCAAGGATGTGGNTGAGAAACGACACA 121
 QY 518 tcgcccctctgaaatcga-cgcaacggaagagctacccctcgtcaaatcggaatccc 576
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 122 TCGCACTCATCAAAATGACCACCGAGGCAAGCNTGCTGTCTGCTGTGGCGGCTCC 181
 QY 577 aaaaattgaaacggcggaatgggtcgctgcacatcgccgcgcccttcggctttgacaac 636
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 182 TCAGAGCTCGGGCGGGAGAGTTCGTGGTGCCTATCGG-AAGCCCGTTTTCCTTCAAAAC 240
 QY 637 agcgtgaccgcccgcgtcgttcgcccaaaaggcagaagc-----ctgcccaac 684
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 241 ACAGTCACCACCGGGATCGTGAGCACCACCCAGAGGGCGGCAAAAGAGCTGGGGCTCCGC 300
 QY 685 gaaagctacacaccccttcattccaaacccgacgtttgcatcaatccgggcaattccggcggc 744
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 301 AACTCAGACATGGACTACATCCAGACCGACGCATCATCAACTATGAAACTCCGGGAGG 360
 QY 745 cc 746
 ||||
 Db 361 CC 362

Search completed: April 14, 2000, 15:42:17
 Job time: 6405 sec

GOGDH
LOCUS GOGDH 2890 bp DNA BCT 15-MAR-1999
DEFINITION G. oxydans gdh gene.
ACCESSION X62710 S60040
VERSION X62710.1 GI:58416
KEYWORDS coenzyme PQQ; gdh gene; gluconic acid production; glucose dehydrogenase; quinoprotein.
SOURCE Gluconobacter oxydans.
ORGANISM Gluconobacter oxydans
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae; Gluconobacter.
REFERENCE 1 (bases 1 to 2890)
AUTHORS Cleton-Jansen, A.M., Dekker, S., van de Putte, P. and Goosen, N.
TITLE A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in Gluconobacter oxydans
JOURNAL Mol. Gen. Genet. 229 (2), 206-212 (1991)
MEDLINE 92017653
REFERENCE 2 (bases 1 to 2890)
AUTHORS Goosen, N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1992) N. Goosen, Dept of Mol Genetics, Gorlaeus Laboratories, Leiden University, P O Box 9502, 2300 RA Leiden, THE NETHERLANDS

FEATURES
source Location/Qualifiers
1. .2890
/organism="Gluconobacter oxydans"
/db_xref="taxon:442"
RBS 242. .245
gene 251. .2677
/gene="gdh"
CDS 251. .2677
/gene="gdh"
/EC_number="1.1.99.17"
/codon_start=1
/transl_table=11
/product="glucose dehydrogenase (pyrroloquinoline-quinone)"
/protein_id="CAA44594.1"
/db_xref="GI:4493391"
/db_xref="SWISS-PROT:P27175"
/translation="MSTISRPRLLWALITAAVFCGAILTVGGAWVAAIGGPLYYVIL
GLALLATAFLSFRNPAALYLFVAVVFGTVIWEITVVGGLDIWALIPRSDIVIIILGIWL
LLPFVSRARSVARGPSCRSFVAVVAVLALFASLFTDPHDSIGELPTQIANASPADPD
NVPASEWHAYGRTQAGDRWSPLNQINATNVSNLKVAWHIHTKMMNSNDPGATNEAT
PIEFNNTLYMCSLHOKLEFVAVGATGNVKNVYDPKLQINPGFQHLTCRGVSFHETPANA
MDSGDNPAPTDCAKDSILPVNDGRLVEVDADTGKTCGFGNGEIDLRVPNQPYTTTPG
QYEPTSPPVITDKLIANSATDNGSVKQASGATQAFDVTGKRVVVFDAENPDNPQL
PDESHPVFHPNSPNSWIVSSYDANLNLVYIPMGVGTDPQWGGDRTKDSEFAPGIVAL
NADTGKLAWFYQTVHDLWDMELPSQPSLVDVTQKDGTLVPAIYAPTGTGDI FVLDRR
TGKEIVPAPETPVPGAAPGDHTSPTQPMSQLTLRPKNPLNDSI WGGTIFDQMFCSI
YFHTLRYEGPFTPPSLKGS LIFPGDLGMFEWGGLAVDPQROVAFANPISLFFVSQVLP
RGPGNLWPEENAKGTGGETGLQHNYGIPYAVNLHPFLDPVLLPFGIKMPCRTPPWGY
VAGIDLKTNKVVWQHRNGTLRDSMYGSSLPPLPIKIGVPSLGGPLSTAGNLGFLTA
SMDYYIRAYNLTKGVWLQDRLPAGAATPITYAINGKQYIVTYAGGHNSFPTRMGDD
IIAYALPDQK"
misc_feature 293. .664
/gene="gdh"
/note="membrane spanning fragment"
mutation 2612
/gene="gdh"
/note="mutation changing substrate specificity"
BASE COUNT 555 a 975 c 786 g 574 t
ORIGIN

Query Match 1.4%; Score 20; DB 1; Length 2890;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 597 atgggtcgctgccatcgcg 616
|||||
Db 340 ATGGGTGCTGCCATCGCG 359

10

11

12

13

14

15

16

Db 64 ACCGACGTTGCCATCA 80
 |||

RESULT 5

V31989/c

ID V31989 standard; DNA; 229 BP.

AC V31989;

DT 25-SEP-1998 (first entry)

DE BS106 polynucleotide clone 1662885 (1).

KW ss; human; BS106; breast tissue gene; breast cancer; detection marker.

OS Homo sapiens.

PN WO9818945-A1.

PD 07-MAY-1998.

PF 31-OCT-1997; U19836.

PR 31-OCT-1996; US-742067.

PA (ABBO) ABBOTT LAB.

PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,

PI Gordon J, Granados EM, Hodges SC, Klass MR, Kratochvil JD,

PI Roberts-Rapp L, Russell JC, Stroupe SD,

DR WPI; 98-272234/24.

PT Detecting BS106 gene products - useful for developing products for

PT detecting, staging, preventing, treating or determining

PS Claim 1: Page 88; 114pp; English.

CC BS106 is a breast tissue gene with which breast cancer and related

CC diseases are associated. The BS106 polynucleotides V31989-V31993 can be

CC detected by BS106 specific polynucleotides or complements acting as

CC useful markers for detection methods. The products and methods can be

CC used for detecting, diagnosing, staging, preventing or treating, or

CC determining predisposition to diseases or conditions of the breast such

CC as breast cancer.

SQ Sequence 229 BP; 43 A; 68 C; 50 G; 62 T;

Query Match

Best Local Similarity 1.2%; Score 17; DB 1; Length 229;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 gggcttcacatcacga 397

|||||

DB 187 GGGCTTCATCATCAGCA 171

RESULT 4

TI9263

ID TI9263 standard; cDNA to mRNA; 196 BP.

AC TI9263;

DT 28-JUN-1996 (first entry)

DE Human gene signature HUMGS00280.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN WO9514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues.

PS Claim 1: Page 345; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in TI9001-726837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 196 BP; 59 A; 55 C; 34 G; 48 T;

Query Match

Best Local Similarity 1.2%; Score 17; DB 1; Length 196;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 709 accgacgttgccatcaa 725

5532



5.5

91

2